A Computational Exploration of Green Tea Catechins and their Interaction with Cardiac Proteins

Botten, Dominic Frith

Awarding institution:
King's College London

A Computational Exploration of Green Tea Catechins and their Interaction with Cardiac Proteins

Dominic Botten

A thesis presented for the degree of
Doctor of Philosophy

Physics Department
King's College London
September 2014

Declaration

This thesis is the result of my own work undertaken between October 2010 and September 2014 at the Physics Department of King's College London, under the supervision of Professor Carla Molteni. I certify that this thesis contains only my own work except where otherwise indicated. This thesis has not been previously submitted in whole or in part for any degree or diploma at this or any other university.

Dominic Botten
September 2014

Acknowledgements

I would like to give heart-felt thanks to my supervisor, Professor Carla Molteni, for her considerable guidance and assistance during this time. In addition I thank my second supervisor Professor Franca Fraternali for her contributions and collaboration on green tea and cardiac troponin. I thank Dr. Ian Robertson for his experimental work and discussions regarding cardiac proteins. I would like to thank my many colleagues at King's who have made it a joy to be here. Special thanks go to Lydia, Giovanni Doni, Giovanni Peralta, Federico, Marco, Zhenwei, Massimo, John and Marcello. I would like to sincerely thank Alessio Comisso for his invaluable help in all matters computational as well as Julia and Paul for their aid throughout my time at King's. Finally I would like to thank my mother for her generosity in providing me with every opportunity I could hope for and her continual support throughout my life.

Abstract

Green tea catechins have been in use for some time as a curative for a number of disorders and illnesses, however their direct benefits are still disputed. In this thesis several computational techniques have been applied to systems involving green tea catechins in order to study their properties and better ascertain their supposed health benefits.

Based on first principles data, classical force fields of four selected green tea catechins (−)-Epigallocatechin 3-Gallate (EGCg), (−)-Epicatechin 3-Gallate (ECg), (−)-Epigallocatechin-3-O-(3-O-Methyl)-Gallate (EGCmg) and (−)-Epigallocatechin (EGC) have been parametrised in order to examine their behaviour under different conditions. Classical Molecular Dynamics coupled with Metadynamics has been used to explore the potential energy surface of these molecules as a function of selected torsions deemed crucial to the conformation of the system. In depth analysis of their interactions under classical conditions has been carried out in gas phase, water and chloroform. The catechins EGCg and EGC were further examined using Density Functional Theory (DFT) to determine their radical properties which are closely linked to their antioxidant activity. By calculating their ionisation potentials, bond dissociation enthalpies and acidity in gas phase, water and benzene solution we find little difference in their energies, despite large structural differences, and we investigate the limit of DFT in accurately describing these systems.

Classical simulations of the cardiac protein Troponin C (a protein known to be crucial to the process of heart contraction), in complex with EGCg, were carried out from the basis of previous experimental work and it was found that EGCg exerts a stabilising effect on the calcium ion mobility present in the system and may help regulate calcium signalling in the cardiac sarcomere. Using sparse experimental data as a guide, alternative protein-ligand binding possibilities were uncovered and used to help validate and expand the current experimental knowledge on the troponin C - troponin I system, giving rise to two low affinity, non-specific binding sites for EGCg which also affect calcium mobility.

Contents

1 Introduction **16**

1.1 Green Tea 16

1.2 Cardiac Function and Regulation 18

1.2.1 Review of Experimental and Computational studies of binding processes involving Troponin C 21

1.3 Thesis Overview 23

2 Theoretical Background **25**

2.1 Chapter Overview 25

2.2 Electronic Structure Methods 25

2.2.1 Hartree-Fock and Møller Plesset Calculation 25

2.2.2 Density Functional Theory 31

2.3 Classical Molecular Dynamics 35

2.3.1 Force Fields for Biological Simulation 37

2.4 Free Energy Exploration and Estimations 41

2.4.1 Metadynamics 41

2.4.2 Poisson-Boltzmann Surface Area Calculations 44

2.4.2.1 Poisson-Boltzmann Equations 46

2.4.2.2 Solvent Accessible Surface Area 49

2.5 Protein - Ligand Docking 49

3 Structural Properties of Green Tea Catechins **52**

3.1 Green Tea Catechins 52

3.2 Force Field Parametrisation for Classical Simulation 53

3.3 Classical Molecular Dynamics of Catechins 59

3.3.1 Catechins in Vacuo 59

3.3.2 Catechins in TIP3P Water and Chloroform 59
3.3.3 Free Energy Conformational Landscape by Metadynamics 69
3.3.3.1 Analysis of the Error in Metadynamics Profiles 71
3.4 Cardiac Drugs: EMD 57033 and Levosimendan 75

4 Electronic Properties of Green Tea Catechins 79
4.1 Chapter Overview 79
4.2 The Importance of Catechin Radicals 79
4.3 Comparison of Radical Scavenging Mechanisms 83
4.4 Spin Density Distributions 87
4.5 Møller Plesset Calculations of Cations 94
4.6 Chapter Conclusions 95

5 Interaction of Epigallocatechin 3-Gallate with Cardiac Troponin C 96
5.1 Troponin C: Function and Mutation 96
5.2 Rigid and Flexible Docking Studies 100
5.2.1 Coordination and Inclusion of Calcium Ions 104
5.2.1.1 Force Field Parameters 105
5.3 Molecular Dynamics Simulations 106
5.3.1 Molecular Dynamics Protocols 106
5.3.2 Secondary Structure Stability 107
5.3.3 Hydrogen Bonding 110
5.3.4 Aromatic Stacking Interactions 113
5.3.5 Comparisons with Experimental NOE Data 115
5.3.6 Calcium Sensitivity 117
5.3.6.1 Calcium Fluctuations 117
5.3.6.2 Calcium Coordination 119
5.3.7 Binding Free Energies 119
5.4 Studying the Conformational Landscape of EGCg in the Hydrophobic Cleft using Metadynamics 122
5.4.1 Dihedral Collective Variables 122
5.4.2 Distance Collective Variables 125
5.5 Chapter Summary 129

6 Interaction of Epigallocatechin 3-Gallate with the Cardiac Troponin C - Troponin I Complex 131

6.1 Chapter Overview 131
6.2 The Troponin C - Troponin I Complex 132
6.2.1 Docking Protocols 133
6.2.2 Molecular Dynamics Protocols and Outcomes 133
6.2.2.1 Protein Stability for Wild-type and Mutated Cardiac Systems 135
6.2.2.2 Calcium Coordination in the presence of cTnI 139
6.2.2.3 Binding Free Energy of the cCTnC·TnI(34-71) Complex 140
6.3 Revisiting Experimental Data: Two Potential Binding Sites . 141
6.3.1 Molecular Dynamics Protocols 142
6.3.1.1 New Force Field Testing 142
6.3.2 Model Preparation and Simulation Protocols 145
6.3.3 Single-Bound EGCg 146
6.3.3.1 Site 1 146
6.3.3.2 Site 2 149
6.3.4 Double Bound EGCg 153
6.3.5 Calcium Mobility and Protein Stability 157
6.3.6 Comparison to Experimental Findings 160
6.4 Chapter Summary 162

7 Conclusions and Perspectives 164

List of Figures

1.1 Camellia Sinensis leaves and the commonly sold green tea extract. Sold as a health supplement, green tea extract is primarily composed of polyphenols (up to 90%) and most of this is in the form of EGCg (around 55%) though content varies by supplier and cost. 18

1.2 A full sarcomere unit shown in relaxed (top) and contracted (bottom) states. The thin filaments are shown as the green helices while the thick filaments with myosin are shown in purple. Z-discs anchor the thin filaments while the M-line maintains the structure of the thick filaments. At full contraction, the thin and thick filaments overlap and the H-band becomes smaller (the total contraction is about 10 nm).[1] . 20

1.3 Generation of a power stroke by Myosin and Actin in Sarcomere Filaments. Major protein structures and sub-units are labelled and intracellular calcium shown. The bending action of the myosin heads is shown by the curved arrow, resulting in the sliding of thick filaments in relation to thin filaments (purple and brown arrows). Some mutations (shown as red spots) of sarcomere proteins enhance contractile function and cause hypertrophic cardiomyopathy or alternatively may reduce the production of contractile force by the sarcomere causing dilated cardiomyopathy.[2] 21

2.1 An example metadynamics simulation. Evolution of the potential of a 1-dimensional system with a 3-minima potential as shown with respect to the collective variable s in the bottom panel. A biased potential V_G is included and its evolution over time is shown by the coloured lines at different stages of the metadynamics simulation. The simulation begins in the middle minima and after 20 Gaussians (blue line) escapes to the second minima (left). After 69 Gaussians (red) this minima is also filled and eventually the system is able to escape into the true minimum (right). After 180 Gaussians (orange) the potential wells have been filled and the metadynamics becomes diffusive. The top panel shows the energy surface reconstructed using these deposited Gaussians.[3] . . 43

3.1 The four catechins selected. Included are labelled rings and torsional angles of interest. 54

3.2 Potential Energy profiles for EGCg as a function of the γ angle for $\theta = 112.5°$ and $\phi = 90°$ calculated with DFT-B3LYP and the classical AMBER force field. The B3LYP profile is in black, the HF in blue, the original profile without torsional correction is shown in red while the corrected profile is shown in orange. The Single point MP2 energies at selected angles are shown in green. The zero of the energy scale is set to the absolute minimum. 56

3.3 Potential Energy profiles for EGCg as a function of the θ angle for $\gamma = 0°$ and $\phi = 90°$ calculated with DFT-B3LYP and the classical AMBER force field. The B3LYP profile is in black, the HF in blue and the classical profile is in orange. The zero of the energy scale is set to the absolute minimum. 56

3.4 Potential Energy profiles for EGCg as a function of the ϕ angle for $\gamma = 0°$ and $\theta = 112.5°$ calculated with DFT-B3LYP and the classical AMBER force field. The B3LYP profile is in black, the HF in blue and the classical profile is in orange. The zero of the energy scale is set to the absolute minimum. 57

3.5 Comparison of the γ torsion potential energy profile of EGCg using B3LYP (black), B3LYP-D (red), and MP2 single point calculations for reference (green). 58
3.6 EGCg with oxygens labelled according to AMBER nomenclature. In ECg O8 is replaced by a single hydrogen while in EGCmg O3 has a methyl group bound to it instead of the hydrogen. In EGC the B′ ring attached to O2 is replaced with a hydrogen. 62
3.7 H-bonds formed in MD simulation with each catechin oxygen. Green: the number of intra-molecular H-bonds formed in vacuo. Blue: number of intra-molecular H-bonds formed in water. Yellow: number of intra-molecular H-bonds formed while in chloroform. Red: number of inter-molecular H-bonds formed in water. Note that ECg lacks O8, while EGC lacks O3, O4, O5 and O11. All intra-molecular bonds are attributed to their acceptor oxygen to prevent double-counting. 63
3.8 Histogram of the major dihedral torsions which determine the structure of EGCg during MD simulation in vacuo (top) and water solution (bottom). 64
3.9 Histogram of the major dihedral torsions which determine the structure of ECg during MD simulation in vacuo (top) and water solution (bottom). 65
3.10 Histogram of the major dihedral torsions which determine the structure of EGCmg during MD simulation in vacuo (top) and water solution (bottom). 65
3.11 Histogram of the ϕ dihedral torsion of EGC during MD simulation in vacuo (top) and water solution (bottom). 66
3.12 Radial Distribution Functions of EGCg oxygens to solvent oxygens. 67
3.13 Radial Distribution Functions of ECg oxygens to solvent oxygens. The ECg lacks the O8 oxygen which changes the characteristics of the O9 and O10 on the B ring. 67

3.14 Radial Distribution Functions of EGCmg oxygens to solvent oxygens. The EGCmg includes a methyl group attached to O3 which distinctly changes the solvation characteristics of the B′ ring. 68
3.15 Radial Distribution Functions of EGC oxygens to solvent oxygens. Note that the EGCg lacks the B′ ring entirely which is replaced with a hydrogen bonded to O2. 68
3.16 Free Energy Surface maps at 300 K as a function of the dihedral collective variables used during metadynamics for EGCg. The top row is the vacuo metadynamics; the middle row is the solvent metadynamics; the bottom row shows the free energy profile as a function of each collective variable integrated over the other two with the black line indicating the solvent profile and the red line indicating the vacuo profile. . 71
3.17 Free Energy Surface maps at 300 K as a function of the dihedral collective variables used during metadynamics for ECg. Contour spacing is 2 kcal/mol. The top row is the vacuo metadynamics; the middle row is the solvent metadynamics; the bottom row shows the free energy profile as a function of each collective variable integrated over the other two with the black line indicating the solvent profile and the red line indicating the vacuo profile. 72
3.18 Free Energy Surface maps at 300 K as a function of the dihedral collective variables used during metadynamics for EGCmg. Contour spacing is 2 kcal/mol. The top row is the vacuo metadynamics; the middle row is the solvent metadynamics; the bottom row shows the single collective variables projected as a function of the free energy with the black line indicating the solvent profile and the red line indicating the vacuo profile. 73
3.19 Free energy profile of the single EGC metadynamics collective variable, the ϕ torsion as a function of the energy. In black is the solvent profile while in red is the vacuo profile. 74

3.20 Overlaid progression of the torsion collective variables for EGCg in water taken at 10,000 Gaussian increments, setting each profile minima to 0 kcal/mol. Particularly noticeable are the final profile (dark green), the first profile (black) and the second profile (red). 74
3.21 Schematic structures of EMD 57033 (top) and Levosimendan (bottom) with labelled rings. 77

4.1 Initial structures of the EGCg neutral and anionic radicals. Left: the B ring interior hydroxyl hydrogen is removed from the 4' carbon position (labelled H7). Right: the exterior hydroxyl hydrogen is removed from the 5' carbon position (labelled H8). Both initial structures were modified by rotating their hydroxyl groups into optimal intra-molecular hydrogen bonding positions before geometric optimisation. Carbon B ring labels are also shown. 84
4.2 Cation spin densities for the EGCg molecule at fixed isovalue 0.005 electrons per Å^3. Left: gas phase optimisations. Right: PCM water solvent optimisations. 88
4.3 Selected neutral H7 EGCg radical spin densities at fixed isovalue 0.005 electrons per Å^3. Left: gas phase optimisations. Right: PCM water solvent optimisations. 89
4.4 Selected neutral H8 EGCg radical spin densities at fixed isovalue 0.005 electrons per Å^3. Left: gas phase optimisations. Right: PCM water solvent optimisations. 89
4.5 Cation spin densities for the EGC molecule at fixed isovalue 0.005 electrons per Å^3. Left: gas phase optimisations. Right: PCM water solvent optimisations. 90
4.6 Selected neutral H7 radical spin densities are shown at fixed isovalue 0.005 electrons per Å^3. Left: gas phase optimisations. Right: PCM water solvent optimisations. 90
4.7 Selected neutral H8 radical spin densities are shown at fixed isovalue 0.005 electrons per Å^3. Left: gas phase optimisations. Right: PCM water solvent optimisations. 91

4.8 Radical optimisations of the EGCg (Left) and EGC (Right) molecules displaying unpaired spin density with B3LYP using benzene implicit solvent at fixed isovalue 0.005 electrons per Å^3. 93
4.9 EGCg and EGC optimised cation with UMP2 single point energy evaluation displaying the MP2 spin density at fixed isovalue 0.005 electrons per Å^3. In this case the EGCg MP2 energy was slightly lower in the ϕ -90° position while for EGC this remained in the +90°. 94

5.1 X-ray crystal structure of the partially resolved core domain of two entwined human cardiac troponin proteins saturated with Ca^{2+}(in pink). Troponin C (shown in Green), troponin T (shown in yellow) and troponin I (shown in red). PDB code 1J1D[4] 98
5.2 View of the dumbbell shaped Troponin C protein sub-unit taken from PDB 1J1D. The N terminal is in light green while the C terminal is in dark green. Three calcium ions are present in the high and low affinity active sites and are in pink.[4] 99
5.3 2KDH pdb stucture showing C terminal troponin C in complex with EGCg. Left: Protein-ligand structure shown with labelled helices and calcium ions. Right: Display of protein molecular surface centred on EGCg. Surface is coloured by residue type where white is hydrophobic, red is acidic, blue is basic and green is polar uncharged. 101
5.4 Clustering results of EGCg to cCTnC hydrophobic cleft using a refined, restricted grid and active residues chosen from NOE data. 102
5.5 Side view of cCTnC with computationally and experimentally docked EGCg ligands. In cyan is the experimental NMR lowest energy structure; in red is M_1; in yellow is M_2 and in purple is M_3. GLY159, which is mutated to an ASP in the mutant models, is displayed in pink while the five experimentally derived NOE residues are in orange. 103

5.6 Initial coordination of the site 1 calcium (pink) with surrounding protein oxygens (red) in cCTnC. Structure taken from pdb 2KDH. 105
5.7 cCTnC with deeply bound EMD 57033 from the lowest energy NMR structure of pdb code 1IH0[5]. For comparison the initial M_{exp} EGCg structure is also shown in purple illustrating the comparative surface binding of EGCg. 108
5.8 Protein backbone RMSD fluctuations over full simulation time for cCTnC models. 109
5.9 EGCg molecule with labelled rings and carbon bonded hydrogen as indexed in AMBER. 111
5.10 cCTnC with labelled aromatic residues which lie on the hydrophobic cleft surface. M_{exp} is also shown for comparison. . 114
5.11 Comparison between experimental NOE data (empty squares)[6] and calculated (diamonds) NOE distances between select hydrogen atoms of EGCg (abscissa) and cCTnC residues (ordinate) in the cCTnC·EGCg models. The colours correspond to the following ranges: medium (2.7 Å - 3.3 Å, purple); weak (3.3 Å - 5.0 Å, red); very weak (5.0 Å - 6.0 Å, orange). . . . 116
5.12 Distributions of the RMSD of calcium ion at site I versus that of site II. Overlaid in black are the wild-type cCTnC simulations while in red are the mutated troponin simulations: (a) M_0, (b) M_{EMD}, (c) M_{exp}, (d) M_1, (e) M_2, (f) M_3. 118
5.13 Top: Free energy surface as a function of the dihedral angles θ, ϕ, γ with the energy scale set to the absolute barrier minimum and contour spacing of 2 kcal/mol. Bottom: Energy profiles of the dihedral torsions (from left to right: θ, ϕ, γ) as collective variables projected over time and setting their absolute minimum to zero. 124

5.14 Close-up of the cCTnC cleft with bound EGCg. Arranged around are coloured representations of the protein residues used as distance collective variables. MET120 is lined in red, LEU136 is lined in orange and VAL160 is lined in blue. Close proximity of the side-chain hydrogens with the relevent carbon bonded hydrogens on EGCg can clearly be seen. . . . 127
5.15 Preliminary free energy profiles of the three distance CVs used in metadynamic simulation. 128
5.16 Reconstructed FES as a function of collective variables for the distance restraint metadynamics simulation with potential walls. 128

6.1 cCTnC·TnI(34-71) with the EGCg ligand computationally docked to the E helix and near the E and H helix interface. The cCTnC remains in green while the cTnI fragment is in red. 134
6.2 Protein backbone RMSDs of the cCTnC·TnI(34-71) complex during simulation broken down by protein sub-unit. In black is the cCTnC while in red is the TnC-(34-71) fragment. On top is replica 1 of the cCTnC·TnI(34-71) simulation while on the bottom is the cCTnC·TnI(34-71)·EGCg replica 1. The drop line shows the point at which the simulations were considered equilibrated and statistics are collected. 137
6.3 MD snapshot after 5 ns of the cCTnC·TnI(34-71)·EGCg replica 1 simulation. Close up view the E - H helix interface displaying several major interactions between EGCg and cCTnC during simulation. In orange are the residues ASP151 (interacting with EGCg ring B here) and TYR150 (π stacking with the B′ ring on helix H). In violet are the GLU94 and GLU95 residues on the E helix with GLU94 currently H-bonding with EGCg ring A. 138
6.4 cCTnC·TnI(34-71) with lowest energy NMR double docked EGCg displaying labelled sites, calcium ions and helix nomenclature. 143

6.5 Close up view of the Site 1 and Site 2 docked locations generated from experiment including the protein molecular surface coloured by residue type: Red - acidic, blue - basic, green - polar un-charged, white - hydrophobic. The two sites share a number of similarities including the presence of hydrophobic residues partially buried in the interface between helices of charged or polar residues. 145
6.6 Replica 1 of the Site 1 simulation at the end of the equilibration phase showing interacting residues. In orange is the TYR150 on the H helix while in violet is the GLU95 and SER98 on the E helix. This is illustrative of the EGCg A ring sliding between the E/H interface and interacting with both sides. 147
6.7 Close up view of the protein-ligand interactions of Replica 1 of the Site 2 simulations after equilibration. In orange are the GLU130 and GLU134 residues on the G helix, in blue is the GLU115 on the F helix and in yellow is the ILE128 residue on the F/G loop. 151
6.8 Protein backbone RMSDs versus simulation time for each model simulated. In each case the black line is replica 1, red is replica 2 and green is replica 3. The drop line at 5 ns indicates where the production runs began. 158
6.9 Calcium ion RMSD (Å) of Site I versus Site II from the production runs of the replica simulations. 159
6.10 Comparison of NOE restraint assignment from experimental docking (hollow black squares) and the computational results. In each graph red squares are for replica 1, blue circles are for replica 2 and green triangles are for replica 3. 161

Chapter 1

Introduction

1.1 Green Tea

Tea, produced from the plant *Camellia Sinensis*, has been cultivated by people since ancient times. Its near mythical discovery by Chinese Emperor Shen Nung led to its inclusion and continuing status as a core part of Chinese medicine up to the present day. In Japan The Way of Tea developed as a cultural ceremony surrounding the preparation, pouring and drinking of green tea. This ceremony is still carried out and is of considerable importance in traditional Japanese culture. The acquisition and production of tea in the 17th and 18th centuries became a contributing factor in the expansionist policies of the colonial European powers of the time (especially Britain), affecting the political landscape to this day. Tea currently stands as the most widely consumed beverage in the world behind water and has attracted significant attention for its believed beneficial effects for a variety of health problems.[7]

Green tea makes up nearly 20% of the total tea produced worldwide and is differentiated from more common black teas by the process of steaming or pan frying shortly after the leaf is picked. This destroys the enzymes naturally present, which would otherwise oxidise the polyphenol contents of the leaf, as well as reducing the water content allowing better storage of the leaves.

It is thought that between 80-100mg of polyphenols are contained in one bag of green tea. Though many catechins are present in small quan-

tities, by far the most abundant of the green tea polyphenols by weight is (−)-Epigallocatechin 3-Gallate or EGCg (roughly 50-75% of total catechin content), a powerful antioxidant which includes a benzenediol ring joined to a tetrahydropyran moiety, a pyrogallol ring and a galloyl ring.[8] These polyphenols will be present in black and oolong tea as well but in lower quantities due to the processing methods involved.

In particular the polyphenolic content of green tea has attracted interest for it purported benefits to combat depression[9–11], influenza[12], HIV[13, 14], obesity[15], Parkinson's and Alzheimer's[16, 17], several forms of cancers[18–21] and cardiovascular diseases[22–24] among others. Considerable interest has recently been placed on green tea anti-cancer activity as well as its ability to inhibit amyloid β fibril formation.[25, 26] EGCg itself has been described as the most efficacious of the green tea catechins and is the primary component of green tea extract, a herbal derivative of green tea shown in Figure 1.1, and has shown promise in the treatment of cardiovascular diseases[22]. The majority of the research attempting to analyse the beneficial effects of green have concentrated on this molecule.

The mechanisms through which green tea catechins affect the body are still not well understood however, and it is instructive to undertake further investigation of these molecules using a range of computational tools and thus gain insights into these molecules' properties. Some studies have shown that the green tea polyphenols could potentially be used to treat cardiovascular diseases, currently the leading cause of death in the developed world[27].

Recently the C terminal of cardiac troponin C (cCTnC, discussed in detail in Section 1.2) has been investigated as a possible target for drugs which modulate the calcium sensitivity of the protein. Experimental studies have shown that the polyphenol EGCg, when bound at the surface of the hydrophobic cleft of the cardiac protein troponin C, may modulate the binding of Ca^{2+} ions to the tissue[28] in a manner similar to that of the known calcium sensitiser EMD 57033. Calcium ions in cardiac tissue regulate heart contraction and several diseases are known to modify this interaction, therefore the biological effect of this polyphenol is of great interest as they could potentially lead to new treatments of heart disease[6, 24, 29].

Figure 1.1: Camellia Sinensis leaves and the commonly sold green tea extract. Sold as a health supplement, green tea extract is primarily composed of polyphenols (up to 90%) and most of this is in the form of EGCg (around 55%) though content varies by supplier and cost.

Mutations on the thin filament are believed to affect the regulation of cardiac muscle contraction including several found in troponin C.[28, 30, 31] Mutations associated with Dilated Cardiomyopathy (DCM) are found to reduce the Ca^{2+} sensitivity of contraction while those associated with Hypertrophic Cardiomyopathy (HCM) increase the Ca^{2+} sensitivity. Current drugs in use as treatment for these diseases do not protect patients from sudden death and so there is a real need to find alternative therapies.

1.2 Cardiac Function and Regulation

Muscle tissue exists in skeletal, cardiac and smooth forms. This work focuses entirely on the aspects of the cardiac muscle system which shares many similarities to the skeletal form including function, structure and contraction regulation. Smooth muscle (as found in the wall of blood vessels for example) is quite different from these other forms and notably lacks the troponin protein. For the remainder of this work only cardiac muscle is assessed. The process of muscle contraction is similar, though not identical, in both skeletal and cardiac muscle forms. In both cases it is carried out by myofibrils within cardiac muscle cells. These myofibrils are in turn formed of long chains of fibrillar units known as sarcomeres. These are the contractile unit of the cell and each sarcomere contains many repeating pro-

tein units primarily composed of a thin filament scaffold of actin with thick filaments of myosin arranged within. The process of muscle contraction is achieved by the movement of the thick and thin filaments with respect to each-other generating tension in the muscle. This is generally described by the sliding filament theory proposed separately by A. Huxley and R. Neidergerke as well as H. Huxley and J. Hanson.[32, 33] The diagram in Figure 1.2 exemplifies this sliding filament model. An action potential is initiated and travels along the surface membrane causing Ca^{2+} ions to enter the cell through calcium channels. This locally increases intracellular Ca^{2+} concentration which leads to a positive feedback loop via further release of Ca^{2+} from the sarcoplasmic reticulum. This calcium induced calcium release generates calcium sparks increasing calcium concentration throughout the cell. Cytoplasmic calcium then binds to active sites on the troponin C protein, signalling a conformational response in the tropomyosin complex which moves from the actin binding site and allows a myosin head group to bind to this position on the actin thin filament. This action is shown in Figure 1.3.

The myosin heads then pull the actin filament toward the centre of the sarcomere causing contraction of the unit. During this time intracellular calcium is being pumped from the cell, dropping the calcium concentration and eventually leading to calcium dissociation from troponin C which causes conformational change to switch the troponin complex back over the myosin binding site on actin and ending the contraction cycle.[34–38]

The Myosin 'cocking' action known as cross-bridge cycling is powered by Adenosine triphosphate (ATP) during the contraction cycle. Once the mysoin head is bound to the thin filament, the phosphate generated in the previous contraction cycle is released resulting in the myosin head rotating toward the center of the sarcomere (the power stroke action occurs here). Following this the attached adenosine diphosphate (ADP) is released. Shortly afterwards a new molecule of ATP attaches to the myosin head, causing the cross-bridge to detach. The myosin head hydrolyzes ATP to ADP and phosphate, which returns the myosin to the cocked position.[1]

A close-up view of the full sarcomere unit is shown in Figure 1.3. Highlighted in the figure are several points where mutations are known to occur

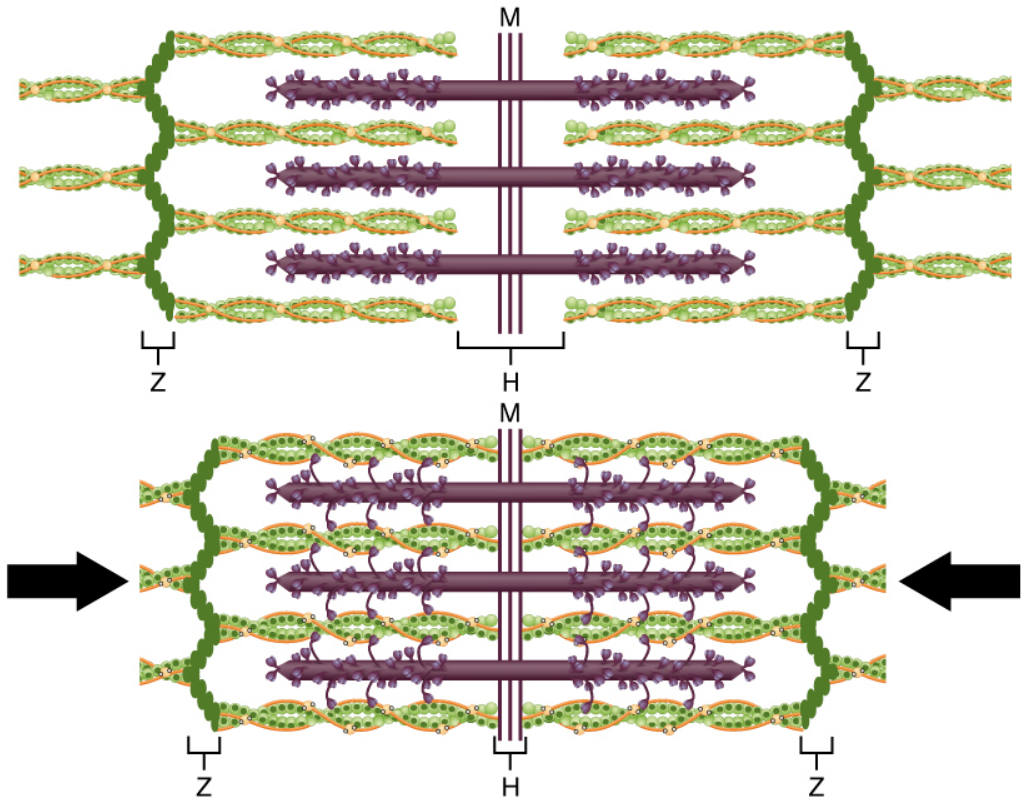

Figure 1.2: A full sarcomere unit shown in relaxed (top) and contracted (bottom) states. The thin filaments are shown as the green helices while the thick filaments with myosin are shown in purple. Z-discs anchor the thin filaments while the M-line maintains the structure of the thick filaments. At full contraction, the thin and thick filaments overlap and the H-band becomes smaller (the total contraction is about 10 nm).[1]

(specifically SER532PRO and PHE764LEU in the myosin heavy chain as well as LYS210 and GLY159ASP in cardiac troponin, though these are not shown in the Figure) which can lead to a reduction of the contractile force and are believed to cause DCM.[2] This weakens the heart, causing it to pump inefficiently and can lead to sudden death in some cases. Alternatively HCM is associated with enlargement of the heart muscle as well as decreased size of the ventricle chambers and is caused by an overload of the contractile force leading to reduced function and mortality.[39, 40]

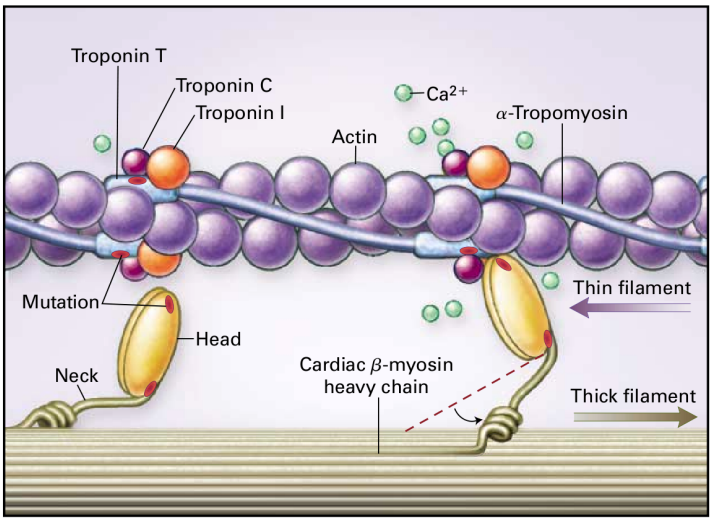

Figure 1.3: Generation of a power stroke by Myosin and Actin in Sarcomere Filaments. Major protein structures and sub-units are labelled and intracellular calcium shown. The bending action of the myosin heads is shown by the curved arrow, resulting in the sliding of thick filaments in relation to thin filaments (purple and brown arrows). Some mutations (shown as red spots) of sarcomere proteins enhance contractile function and cause hypertrophic cardiomyopathy or alternatively may reduce the production of contractile force by the sarcomere causing dilated cardiomyopathy.[2]

1.2.1 Review of Experimental and Computational studies of binding processes involving Troponin C

A number of studies have investigated the binding properties of ligands to troponin C. These have primarily targetted the N terminal domain, though some research has suggested that several molecules also bind to the C terminal. Examples are listed in the following.

Levosimendan: This calcium sensitizer has been extensively studied in connection with troponin C and is currently used as a treatment for patients with severe decompensated heart failure.[41] Levosimendan is believed to increase Ca^{2+} sensitivity by binding directly to the N domain of troponin C[42] though it has been shown that it is also capable of binding to the C domain.[43] NMR chemical shift mapping was used to demonstrate binding to cNTnC in the presence of $cTnI_{32-79}$ (the N terminal) and $cTnI_{128-180}$ (the switch region). More recently 2D HSQC NMR spectroscopy has been

used to compare this binding between Levosimendan and several of its analogs. Among these, only those with a specific chiral methyl group were found to induce chemical shifts and so it was concluded that this could be the key pharmacophore responsible for binding pyridazinone derivatives to cNTnC.[44]

Bepridil: This Ca^{2+} channel blocker was found to bind to cNTnC in the presence of cTnI and alone, though with reduced affinity when cTnI is present.[45, 46] NMR studies as well as stop flow fluorescence techniques have concluded that two of the three potential binding sites for Bepridil are blocked by cTnI though even with only one site occupied Bepridil is believed to slow the release of Ca^{2+} from site II of cNTnC.[47] Computational molecular dynamics on the complex concluded that competition between Bepridil and cTnI was due to the ligand binding to the hydrophobic portion of the N terminal of troponin C.[48] Computational screening and molecular dynamics have been used to compare a number of molecular analogues to Bepridil, several of which were found to have improved binding chraracteristics.[49]

Trifluoperazine: NMR spectroscopy and chemical shift mapping have been used to show that this molecule binds to TnC in 3-4 places, two of which are within the hydrophobic binding pocket of the N domain.[50, 51] Binding to these sites appears to have a similar effect to Bepridil and increases the Ca^{2+} affinity of TnC.[46, 52]

EMD 57033: This drug has been found to increase Ca^{2+} sensitivity in a stero-specific manner (its (–)-enantiomer, EMD-57439, did not exhibit the same behaviour).[53] Unlike the previous drugs, EMD 57033 was found to bind to the C domain of troponin C using direct binding measurements[54] and NMR shift mapping.[55] The NMR work indicated that the sensitizing ability derives from a modification of the interaction between cCTnC and cTnI by competing with $cTnI_{34-71}$ thereby enhancing the binding of C terminal cTnI to cTnC.[5]

EGCg: This green tea polyphenol has been investigated by the group of Sykes et al.[6] for its binding characteristics to troponin C. The structure of the protein-ligand complex was evaluated using NMR spectroscopy and a solution NMR structure was deposited in the protein data bank. Like EMD 57033, this molecule was found to preferentially bind to the C terminal of

cardiac troponin C. For isolated cCTnC, the binding was found to be within the hydrophobic cleft of the protein (see Chapter 5). This cleft is usually the location for binding of the $cTnI_{34-71}$ fragment; however the group were unable to precisely determine the binding of EGCg to the protein complex when this fragment was included. The binding of EGCg to troponin C is the subject of Chapter 5 and 6.

Experimental studies on the troponin protein have also focussed on the binding characteristics of calcium ions to the N terminal of troponin C, though some work on the high affinity calcium sites on the C terminal have also been undertaken.[56–60] This has lead to a greater understanding of the processes underlying contraction-relaxation mechanisms in cardiac and skeletal muscle proteins. Investigations have also shown that magnesium ions (Mg^{2+}) will competitively bind to troponin C in the same calcium binding sites and that the increase of magnesium ions in the sarcomere reduces the affinity of the troponin protein for calcium. In addition, minor changes in pH (from 7.2 to 6.5) were shown to decrease calcium affinity for troponin C.[61]

Several studies have evaluated the relationships between the protein sub-units of troponin, including computational docking of the troponin I to troponin C[62] as well as experimental interactions of the troponin I-troponin C complex.[63] The interaction of all three sub-units (TnC, TnI and TnT) were investigated using molecular dynamics (MD) simulations for both skeletal and cardiac troponin to monitor the cooperativity between the units and to better understand the effects of mutation on the protein.[64] Recent work has also been undertaken to precisely determine the position of the troponin protein with respect to the tropomyosin-actin complex by reconstructing thin filaments through imaging the proteins with transmission electron microscopy and then post-processing these images using single particle analysis to build up a more complete picture of the complex.[65]

1.3 Thesis Overview

This thesis offers a detailed analysis of structural and electronic features of a selection of green tea catechins in gas phase, solution and in complex with

cardiac troponin C calculated through computational methods. Chapter 2 details the theory behind the methodology used, including *ab initio* and classical simulation techniques as well as free energy methods which have been used throughout. Chapter 3 develops a force field parametrisation for green tea catechins to be used in classical molecular dynamics simulation and gives discussion of the structural properties of these molecule in vacuo, water and chloroform. Chapter 4 discusses the electronic structure properties of the green tea catechins, primarily EGCg, in the context of their antioxidant behaviours. Analysis of the anionic, cationic and several neutral radicals is untertaken here. Chapter 5 evaluates the binding of EGCg to the cardiac muscle protein Troponin C. Simulations of this molecular complex are compared with the available experimental results. Free energy data is gathered for the binding of the system as well as production of free energy conformational maps from metadynamics. Chapter 6 investigates the binding of EGCg to troponin C in the presence of the anchoring part of troponin I (which occupies the hydrophobic cleft of troponin C seen to be the binding site of EGCg in Chapter 5). Finally Conclusions and future outlooks based on the work presented are considered.

Chapter 2

Theoretical Background

2.1 Chapter Overview

This Chapter describes the computational techniques used in the following work. Simulation methods provide a powerful tool to tackle problems from the atomic to the macromolecular scale. Initially electronic structure methods are outlined as a means of solving the many-body Schrödinger equation. Following this, schemes of free energy evaluation are also discussed which can be used to contribute important information on the chemical and biological processes of interest. Finally a brief description of computational protein-ligand docking methods is undertaken with a view to the types of binding calculations carried out on biological systems.

2.2 Electronic Structure Methods

2.2.1 Hartree-Fock and Møller Plesset Calculation

Much of electronic structure theory is built upon the fundamental Hartree-Fock theory and, while more complex and accurate methods have been formulated to solve the many-body problem, it remains very useful in a variety of systems and contexts.[66–68] At its heart, Hartree-Fock theory seeks to solve the many-body electronic Schrödinger equation obtained by invoking the Born-Oppenheimer (BO) approximation, which states that, because the nuclei in a molecular system are considerably more massive

than the electrons but experience a similar force, the electrons will effectively react instantaneously to the motion of the nuclei. We then consider that, for any nuclear configuration, the electrons will be in the ground state of the system. This leaves us with the electronic time independent Schrödinger equation for a given nuclear configuration:

$$\hat{H}_{el}\Psi = \left[\hat{T} + \hat{V} + \hat{U}\right]\Psi = E\Psi \tag{2.1}$$

used to solve the ground state wavefunction Ψ. Here $\hat{H}$ is the Hamiltonian, $\hat{T}$ is the kinetic energy, $\hat{V}$ is the external potential exerted on the system from the nuclei, $\hat{U}$ is the electron-electron interaction energy and E is the energy of the system. The energy E here gives the potential energy experienced by the nuclei and enables us to find important aspects of the system which depend on the nuclear configuration such as equilibrium geometries and vibrational frequencies. Even with the BO simplifying assumption, this equation is prohibitively expensive to solve for the many electron wavefunction and so several methods were introduced to deal with this. One of the earliest was the Self-Consistent Field method (SCF) or Hartree-Fock method. The initial assumption to the theory was that electrons interacting with each other can be simplified by assuming that each electron interacts only with a mean field which describes the interactions with all the other electrons. We can then produce the Hartree Ansatz:

$$\Psi_{HA}(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N) = \phi_1(\mathbf{r}_1)\phi_2(\mathbf{r}_2)\ldots\phi_N(\mathbf{r}_N) \tag{2.2}$$

where $\phi(\mathbf{r})$ are orbitals describing the electrons of the system. This however does not satisfy the antisymmetric principle for fermions, which states that a wavefunction describing fermions should be antisymmetric with respect to the exchange of any set of space-spin coordinates, required by the Pauli exclusion principle. So the wavefunction is re-written in the form of a Slater Determinant of spin-orbitals:

$$\Psi = \frac{1}{\sqrt{N!}} \begin{vmatrix} \chi_1(\mathbf{x}_1) & \chi_2(\mathbf{x}_1) & \dots & \chi_N(\mathbf{x}_1) \\ \chi_1(\mathbf{x}_2) & \chi_2(\mathbf{x}_2) & \dots & \chi_N(\mathbf{x}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \chi_1(\mathbf{x}_N) & \chi_2(\mathbf{x}_N) & \dots & \chi_N(\mathbf{x}_N) \end{vmatrix} \tag{2.3}$$

where N is the total number of electrons in the system and the spin orbital χ accounts for both the spatial and spin degrees of freedom ($\underline{x}$). Using a Slater determinant as a trial wavefunction implies that the electrons are moving independently of the others, feeling only the Coulomb repulsion of the average of the other electrons and an exchange interaction due to the antisymmetrisation. Using the variation method we can derive the Hartree-Fock equations by setting the condition that the derivative of the expectation value of the Hamiltonian, δE, is zero. Hence the optimal molecular spin-orbitals are those which minimise $E = \langle\Psi|\hat{H}_{el}|\Psi\rangle$. This can be effectively achieved using Lagrange multipliers with the constraint that the molecular orbitals remain orthonormal:

$$\delta E - 2\delta \sum_i \sum_j \varepsilon_{ij} S_{ij} = 0 \tag{2.4}$$

where we have used the convention that the multiplier λ_{ij} is written $-2\varepsilon_{ij}$ as it is related to the molecular orbital energies. S_{ij} is the overlap integral between two orbitals and ensures the orthonormality condition. It is possible to manipulate the resulting equations such that the Lagrange Multipliers are zero when $i \neq j$ so that the Hartree Fock equations take the standard eigenvalue form:

$$\left[h(\mathbf{x}_1) + \sum_{j \neq i} \mathcal{J}_j(\mathbf{x}_1) - \sum_{j \neq i} \mathcal{K}_j(\mathbf{x}_1) \right] \chi_i(\mathbf{x}_1) = \varepsilon_i \chi_i(\mathbf{x}_1) \tag{2.5}$$

where

$$h_i = -\frac{1}{2}\nabla_i^2 - \sum_A \frac{Z_A}{r_{iA}} \tag{2.6}$$

is the 'core' one electron operator, ε_i is the energy eigenvalue of orbital

χ_i and Z_A the charge of nucleus A,

$$\mathcal{J}_j(\mathbf{x}_1) = \int d\mathbf{x}_2 |\chi_j(\mathbf{x}_2)|^2 r_{12}^{-1} \tag{2.7}$$

is the Coulomb operator giving the average local potential at point $\mathbf{x}_1$ from the charge distribution from the electron in orbital χ_j and

$$\mathcal{K}_j(\mathbf{x}_1)\chi_i(\mathbf{x}_1) = \left[\int d\mathbf{x}_2 \chi_j^*(\mathbf{x}_2)\chi_i(\mathbf{x}_2) r_{12}^{-1}\right] \chi_j(\mathbf{x}_1) \tag{2.8}$$

is known as the exchange operator and is defined in terms of its effect in acting upon the spin orbital χ_i. We can make use of the fact that $[\mathcal{J}_i(\mathbf{x}_1) - \mathcal{K}_i(\mathbf{x}_1)]\,\chi_i(\mathbf{x}_1) = 0$ to introduce the Fock operator:

$$f(\mathbf{x}_1) = h(\mathbf{x}_1) + \sum_j \mathcal{J}_j(\mathbf{x}_1) - \mathcal{K}_j(\mathbf{x}_1) \tag{2.9}$$

which reduces Equation 2.5 to:

$$f(\mathbf{x}_1)\chi_i(\mathbf{x}_1) = \varepsilon_i \chi_i(\mathbf{x}_1) \tag{2.10}$$

In practice, for a closed shell system, by writing the spatial part of the spin-orbitals as linear combinations of a finite set of atomic orbitals or LCAO (also known as basis functions)

$$\phi_i = \sum_{\mu=1}^{K} c_{\mu i}\phi_\mu \tag{2.11}$$

where ϕ_μ is one of K atomic orbitals and $c_{\mu i}$ is a coefficient, one can obtain Roothan-Hall equations for the Fock matrix:

$$\sum_\nu F_{\mu\nu} C_{\nu i} = \epsilon_i \sum_\nu S_{\mu\nu} C_{\nu i} \tag{2.12}$$

where ϵ_i are the orbital energies, $S_{\mu\nu} = \int d\mathbf{x}_1 \phi_\mu^*(\mathbf{x}_1)\phi_\nu(\mathbf{x}_1)$ is the overlap matrix and $F_{\mu\nu} = \int d\mathbf{x}_1 \phi_\mu^*(\mathbf{x}_1) f(\mathbf{x}_1)\phi_\nu(\mathbf{x}_1)$ is the Fock matrix.[66]

Equation 2.12 requires an iterative method to obtain the molecular orbitals and so is often called the Self-Consistent-field (SCF) procedure. Several types of basis sets have been proposed for these calculations with the

most popular being the Slater-type or Gaussian-type. Slater functions are an obvious choice but become difficult to evaluate for molecular orbitals and so the Gaussian-type is often preferred, which have the advantage that the product of two Gaussians can be expressed as a single Gaussian located at the lines joining the two. This can make the evaluation of the combination of orbitals much more efficient; however replacing a Slater type orbital with a single Gaussian orbital will badly underestimate the cusp at the origin as well as decaying faster than we would like, equivalent to underestimating the long-range atom overlap and the charge and spin density at the nucleus. Hence we use a linear combination of Gaussian functions to more closely approximate the Slater-orbital, though this does not fully correct the problem. A possible improvement was found by splitting the treatment of core and valence orbitals into seperate Gaussian orbitals. This works as the core orbitals do not affect chemical properties much and only slightly change from one molecule to another. Additional basis functions have been proposed to ameliorate this issue, including polarisation functions which have a higher angular momentum quantum number and so, for example, give a slight p orbital character to hydrogen as well as d orbitals to the first and second row elements. Where molecular species have significant electron density far from the nuclei, such as anions or lone pairs, these basis sets are still found to be deficient. To address these problems diffuse basis sets were introduced which contain an additional set of diffuse s-type and p-type Gaussian functions.

For open-shelled systems, such as radicals with one or more unpaired electron, the Roothan-Hall equations are not applicable. Instead approaches known as the Restricted Hartree-Fock (RHF) and Unrestricted Hartree-Fock (UHF) have been developed. The RHF theory uses doubly occupied molecular orbitals as far as possible, then uses singly occupied orbitals afterward. The alternative UHF method uses two sets of molecular orbitals, one for all electrons of spin up and the other for spin down. This entails two Fock matrices, one for each spin type, as well as two density matrices where the full matrix is the sum of these. The UHF version is the more general form and has shown to be the better method for dealing with atoms close to their dissociation limit for example; however, as the spin-orbitals are not explicitly

paired, 'spin density', a property much like charge density, will exist from the excess of electron spin in the system. The UHF method can also suffer from a problem known as spin contamination in systems where some higher spin states become mixed in with the correct one.[69] This arises from the fact that the UHF wavefunction is not an eigenfunction of the total spin, $\langle S^2 \rangle$, unlike in the RHF case. Unfortunately this error is not systematic and so can be difficult to deal with. Usually spin annihilation is performed to reduce the overall contamination to an acceptable level (within 10% of the true spin is usually considered reasonable for organic systems) though other methods are available and none completely remove this problem.[67]

A significant drawback to HF theory is the problem of electron correlation. In reality electron motion is correlated and they will tend to avoid each other; however as HF is a mean field theory the electrons are assumed not to interact with neighbouring electrons but with the average field of the whole system. This leads to HF theory giving higher energies than should be the case and the correlation energy is thus defined as the difference between the HF energy and the exact energy. A number of methods have been implemented to attempt to resolve this issue, including the many body perturbation theory as proposed by Møller and Plesset and commonly known as Møller Plesset, or MP, methods.[70] Here the Hamiltonian operator is expressed as a combination of a 'zeroth order' Hamiltonian ($\mathcal{H}_0$), our original Hamiltonian from above, and a perturbation ($\mathcal{V}$). In order to improve the eigenfunction and eigenvalues of $\mathcal{H}_0$ we write:

$$\mathcal{H} = \mathcal{H}_0 + \lambda \mathcal{V} \tag{2.13}$$

where λ is a parameter which varies between 0 and 1, one being the point that we achieve the 'true' Hamiltonian value. Eigenvalues of the wavefunction, Ψ_i, and energies E_i are then written:

$$\Psi_i = \Psi_i^{(0)} + \lambda \Psi_i^{(1)} + \lambda^2 \Psi_i^{(2)} \ldots = \sum_{n=0}^{\infty} \lambda^n \Psi_i^{(n)} \tag{2.14}$$

$$E_i = E_i^{(0)} + \lambda E_i^{(1)} + \lambda^2 E_i^{(2)} \ldots = \sum_{n=0}^{\infty} \lambda^n E_i^{(n)} \tag{2.15}$$

where the superscript denotes the correction to the specific order, for example $E^{(2)}$ is the energy correction to second order, and the summation is truncated to the desired order. The perturbation is related to the fact that the true Hamiltonian is given by the sum of nuclear attraction and electron repulsion terms and is thus written as:

$$\mathcal{V} = \sum_{i=1}^{N} \sum_{j=i+1}^{N} \frac{1}{r_{ij}} - \sum_{j=1}^{N} (f_j + \mathcal{K}_j) \tag{2.16}$$

The sum of zeroth order and first order energies will correspond to the HF energy and so the correction at least to second order must be used and is often written as MP2. This is a popular way to account for electron correlation effects however it is possible the final energy will be lower than the true energy as the theory is not variational. It is also a highly computationally expensive method and so is often restricted to single point energy evaluations at geometries calculated with lower levels of theory.

2.2.2 Density Functional Theory

One of the most well known and successful approaches to the electronic structure problem is Density Functional Theory (DFT) which has been widely used in solid state and condensed matter physics, but also for molecular systems. The theory attempts to address inaccuracies with HF, while reducing the computational cost relative to other post-HF methods.[71] Despite the advent of fast, parallel computing allowing for larger systems to be calculated, post-HF methods are still unsuitable for complex systems (though the software CP2K has shown itself capable of running post-HF methods on large systems[72]) as they are still very expensive to calculate numerically and HF is unable to deal with electron correlation effects (a consequence of using the mean field approximation). Density Functional Theory is a ground state many body theory using the electronic density, $\rho(\mathbf{r})$, as the quantity which determines the state of the system:

$$\rho(r_1) = N \int dr_2 \dots dr_N \Psi^*(r_1, \dots r_N) \Psi(r_1, \dots r_N) \tag{2.17}$$

This is part of the Hohenburg-Kohn theorems[73] which shows that the

ground-state properties of a system can be expressed as functionals of the electron density and that the optimised ground state of the system can be found with the proper choice of this density. In this manner it is possible to write the ground state energy as

$$E = \min_{\rho}[E[\rho(\mathbf{r})]] = \min_{\Psi}\left[\left\langle \Psi \left| \hat{H} \right| \Psi \right\rangle\right] = \min_{\rho}\left[F[\rho(\mathbf{r})] + \int \rho(\mathbf{r})V(\mathbf{r})d\mathbf{r}\right] \tag{2.18}$$

Here $F[\rho(\mathbf{r})]$ is a universal function of the charge density which is not known.

This was later extended into the popular Kohn-Sham approach for DFT which introduces a fictitious set of electrons that are non-interacting but provide the same electron density as that of interacting electrons.[74] By constructing a set of one electron orthonormal orbitals the electron density at a point $\mathbf{r}$ can be written as a sum of the occupied orbitals squared:

$$\rho(\mathbf{r}) = 2\sum_{i=1}^{N_{occ}} |\psi_i(\mathbf{r})|^2 \tag{2.19}$$

where ρ is the electron density, ψ_i is the Kohn-Sham orbital and all states are assumed to be doubly occupied.[66] The KS orbitals are used as solutions to the one electron Kohn-Sham Equations (Equation 2.21) below. In modern calculations the atomic orbitals are commonly expressed as basis sets built up of Slater-type or Gaussian-type orbitals (as well as plane wave methods). A brief discussion on these orbitals is given in Subsection 2.2.1. The energy of the system can then be expressed in terms of this density:

$$\begin{aligned} E(\rho) = 2\sum \int \psi_i^* \left(-\frac{\nabla^2}{2}\right)\psi_i d\mathbf{r} + \int V\rho(\mathbf{r})d\mathbf{r} \\ + \frac{1}{2}\int\int d\mathbf{r}d\mathbf{r}' \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} + E_{XC}[\rho(\mathbf{r})] \end{aligned} \tag{2.20}$$

where ∇ is the gradient operator, $d\mathbf{r}$ indicates the integration over all spatial coordinates and V is the static potential derived from the nuclear potential. The first term is the kinetic energy of the non-interacting electrons, the second term is the Hartree energy describing the electrostatic

interaction between clouds of charge and the exchange-correlation energy E_{XC} contains all the other terms. The minimisation of the energy functional with respect to the density leads to the one-electron Kohn-Sham equations

$$\left\{-\frac{1}{2}\nabla^2 + \int \frac{\rho(\mathbf{r})d\mathbf{r}}{|r-r'|} + V + \frac{\delta E_{XC}}{\delta\rho(\mathbf{r})}\right\}\psi_i = \varepsilon_i\psi_i \tag{2.21}$$

which are solved self consistently.

One of the simplest methods of including the exchange-correlation energy is the local density approximation (LDA) which assumes that charge density varies slowly in a system and so a localised region may be treated as a uniform electron gas (for which the exchange-correlation can be calculated). Thus if ε_{XC} is the exchange-correlation energy per particle in the uniform electron gas then:

$$E_{XC}[\rho(\mathbf{r})] \cong \int \rho(\mathbf{r})\varepsilon_{XC}[\rho(\mathbf{r})]d\mathbf{r} \tag{2.22}$$

This is acceptable for single atoms (and works well for some solids) but for molecules (where LDA badly overestimates bond strengths) a gradient correction is usually required. This changes the exchange-correlation functional to a function both of the local density and of the local gradient of the density:

$$E_{XC}[\rho(\mathbf{r})] = \int \varepsilon_{XC}(\rho(\mathbf{r}), |\nabla\rho(\mathbf{r})|)\rho(\mathbf{r})d\mathbf{r} \tag{2.23}$$

which improve the calculation of binding and bond energies at a relatively modest increase in computational cost. Common gradient corrections include those in the 1996 gradient-corrected correlation functional of Perdew, Burke and Ernzerhof (PBE)[75] and that of Becke[76] as well as a correlation functional such as the Lee-Yang-Parr (LYP)[77] functional. The so called BLYP functional has shown to be robust in describing many organic molecules and is a popular choice for biological systems. Over time, several additive terms have been proposed as well as alterations to the functionals used such as the hybrid functionals PBE0[78] and Becke3[79] which mix in, using a three parameter set, exact exchange from HF theory with the Becke 88 (B88) exchange functional and correlation from LYP, also called

B3LYP. These hybrid functionals have shown to be very successful for a range of systems and remain the most popular in use.

Pseudopotentials can be introduced to further simplify the problem which separate the core and valence electron wavefunctions. As the core electrons are tightly bound to the nucleus they play little role in atomic interactions and so pseudopotentials are used to reproduce their effect while attempting to accurately model the valence electrons. This reduces the number of electrons to be explicitly modelled and can reduce the number of basis sets required for an accurate calculation of the system. There are many different pseudopotentials available depending on the system to be modelled and the accuracy required and are particularly useful when using a plane wave basis set to expand the orbitals.[66, 80]

One problem with DFT which arises is the so-called self-interaction where the interaction of an electron with the field it generates does not cancel out exactly (unlike in the HF case). Several self-interaction corrections (SIC) have been proposed but these are not applicable to all systems.[81]

Longe-range corrections have also been included in hybrid functionals such as CAM-B3LYP to better produce the effect of charge transfer and other long range properties using a Coulomb attenuating method. In this case, at the short range exchange mixing is 0.19 HF and 0.81 B88 while at longer ranges a larger mixing of HF is used (0.65 HF to 0.35 B88). To smooth the region between these ranges, an intermediate region uses a smooth error function with a parameter 0.33.[82]

A further continuing issue in this field is the treatment of London dispersion forces. Dispersion interactions play an important role in many chemical processes. These weak interactions are due to instantaneous dipoles which arise during the fluctuations in the electron clouds and can induce a dipole in neighbouring atoms, giving rise to an attractive effect. They are known to be important in the study of noble gases, molecules interacting with substrates or biomolecules, for example, where weak interactions can be significant.[66]

Recently, several methods of introducing dispersion corrections have been proposed including using DFT functionals parameterised to produce dispersion behaviour[83], functionals with dispersion correcting potentials[84,

85] as well as functionals corrected by empirical coefficients.[86–88]

Previous functionals have been shown to badly account for this interaction leading to poor results in some circumstances.[89, 90] Many attempted empirical solutions to this problem have led to much debate as to whether the fitting parameters are suitable for multiple systems. They can also lead to over-correction in some cases so the search for new, better functionals is an on-going topic.

Uncorrected functionals can still perform very well compared to experimental results in many areas.[91] Where systems are dominated by electrostatic interactions, for example, a dispersion correction can have very small effects. In addition, uncorrected functionals such as PBE0 and B3LYP have been very well tested over the years on a wide variety of systems, and this allows a level of comparability which newer functionals necessarily lack. Nevertheless, the acceptable computational increase in using these corrections combined with the improvement they provide means that dispersion corrected functionals are likely to be commonly used and improved upon as time passes.

2.3 Classical Molecular Dynamics

Thus far the methods described have primarily revolved around time independent calculations of a system in its ground state. While it is undeniable that this provides crucial information about the system of interest, the truth remains that most or all systems we would like to investigate are dynamically changing and thus we need to turn to methods such as the well known Molecular Dynamics (MD) in order to further our scientific goals. Here the dynamics of a system of classical particles is evaluated by integrating Newton's equations of motion:

$$\frac{d^2\mathbf{r}_i}{dt^2} = \frac{\mathbf{F}_i}{m_i} \tag{2.24}$$

which describes the evolution of the position $\mathbf{r}_i$ of a particle i with mass m_i subject to the force $\mathbf{F}_i$. We can assume in most cases that the nuclei move as classical particles in the electronic ground state (either calculated via

quantum mechanical methods like DFT or as empirical potentials). This integration is carried out in a discrete manner from time t to time $t + \Delta t$ where Δt is a time-step determined by the smallest motion of the system (usually the vibration of bonds containing hydrogen) in order to prevent problems such as energy drift. This allows average properties of the dynamic system to be calculated. The velocity Verlet algorithm is popular and often used for these purposes and takes the form of:

$$\mathbf{r}_i(t + \Delta t) = \mathbf{r}_i(t) + \Delta t \mathbf{v}_i(t) + \frac{1}{2}\Delta t^2 \mathbf{a}_i(t) \tag{2.25}$$

$$\mathbf{v}_i(t + \Delta t) = \mathbf{v}_i(t) + \frac{1}{2}\Delta t[\mathbf{a}_i(t) + \mathbf{a}_i(t + \Delta t)] \tag{2.26}$$

where $\mathbf{r}_i$, $\mathbf{v}_i$ and $\mathbf{a}_i$ are the position, velocity and acceleration of particle i. There are a number of implementations which can increase the speed of these integrator methods such as constraining the bonds involving hydrogen in the system (as with the SHAKE algorithm[92]) allowing larger time-steps to be used. Thermostats and barostats can be added depending on the type of ensemble to be reproduced.[66, 93]

While *ab initio* quantum mechanical calculations can be used in MD simulations, most bio-molecular systems are too large for them to be effectively used due to the sheer computational cost. This is because the electronic wavefunction must be taken into account which makes the calculations simply too costly. Classical Molecular Dynamics, whereby forces are derived by empirical potentials (force fields), can be feasibly used to investigate time scales of up to 100-1000's of nano seconds at the cost of some accuracy and can provide some insights into the behaviour of systems with thousands of atoms. These force fields express the potential energy of the system as a function of distances, angles, torsions, etc. and include a set of parameters (containing the force constants, equilibrium properties of bond lengths etc.) to be used in the calculations. These parameters are built up from experimental data or from quantum mechanical calculations and are often refined by further work over time.[93] Using these force fields as a framework, methods such as Molecular Dynamics become relatively simple and less computationally expensive for large systems.

2.3.1 Force Fields for Biological Simulation

Simple force fields used today for (bio)-molecular systems can be reduced to components of the inter- and intra-molecular forces, which can be described by the bonding terms (stretching, bending and bond rotations or torsions) and by the non-bonded interactions (van der Waals and electrostatic). These terms define the potential energy of the system and the forces on the atoms are then simply calculated as the negative gradient of this potential.[66, 94] An example of such a force field is shown in Equation 2.27.

$$V(r^N) = \sum_{bonds} \frac{1}{2} k_b (l - l_0)^2 + \sum_{angles} \frac{1}{2} k_a (\theta - \theta_0)^2 + \sum_{torsions} \frac{1}{2} V_n [1 + cos(n\omega - \gamma)] + \sum_{j=1}^{N-1} \sum_{i=j+1}^{N} \left\{ \epsilon_{i,j} \left[\left(\frac{r_{0ij}}{r_{ij}} \right)^{12} - \left(\frac{r_{0ij}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \right\} \quad (2.27)$$

which is the generalised form of the AMBER field[95] which has been widely tested and refined for use with biological systems and is the force field used in this work. Here the terms k_b and k_a are constants of the stretching and bending terms respectively, l_0 and θ_0 are equilibrium positions of the bond length and bond angle respectively. The torsional term contains V_n, related to the potential barriers to rotation of a bond, n is the multiplicity of the minima points around the bond, ω is the torsional angle and γ is the offset related to this. The non-bonded terms relate to the Lennard-Jones and Coulomb potential, N is the total number of particles, $\epsilon_{i,j}$ is the minimum potential energy, r_{0ij} is the "diameter" of a particle, q_i is the charge on particle i and r_{ij} is the distance between particles.

The bonded part is short-ranged and therefore relatively easy to calculate. This means the most computationally intensive part of classical MD then becomes calculating the non-bonded term in Equation 2.27 which can include very many long range interactions in large systems as each particle is assumed to interact with each other particle in the simulation box and with each particle image in an infinite array of periodic cells (when using period boundary conditions). This can be simplified using an Ewald summation where interactions are split into short and long ranges. The point

charges in the system are assumed to be a set of screened charges with Gaussian distributions, then short range interactions are summed in real space (within a cutoff distance) while the long range interactions (which describe the interactions of the unit cell with those of the repeated images) are summed in Fourier space. These contributions are then given by:

$$U_{long-range} = \frac{1}{2V} \sum_{k \neq 0} \frac{4\pi}{k^2} |\rho(\mathbf{k})|^2 exp(-k^2/4\alpha) \tag{2.28}$$

$$U_{short-range} = \frac{1}{2} \sum_{i \neq j}^{N} q_i q_j erfc(\sqrt{\alpha} r_{ij})/r_{ij} \tag{2.29}$$

$$U_{corr} = (\alpha/\pi)^{\frac{1}{2}} \sum_{i=1}^{N} q_i^2 \tag{2.30}$$

Where U_{corr} corrects the self interaction term which appears in Equation 2.28. In these equations V is volume, $\mathbf{k} = (2\pi/L)\mathbf{l}$ with $\mathbf{l} = (l_x, l_y, l_z)$ being the lattice vectors in Fourier space, α determines the width of the Gaussian charge distribution, $q_{i/j}$ is the charge on particle i/j respectively and $erf(x) \equiv (2/\sqrt{\pi}) \int_0^x exp(-u^2)du$ is the error function with $erfc(x) \equiv 1 - erf(x)$ and u represents the tail part of the potential. These sums converge relatively quickly and so may be truncated without great loss in accuracy while improving the computational time requirements.[66, 93, 96] The efficiency of this calculation in terms of CPU time can be further improved by the use of a mesh or grid which the charges of the system are placed upon. The long-range interactions are then solved using the Fast Fourier transform while the short-range interactions continue to be calculated directly from particle-particle interactions.[93]

Molecular dynamics makes use of the ergodic hypothesis which states that all accessible microstates are equally probable over a long period of time allowing macroscopic properties of the system to be calculated. A molecular dynamics simulation is not usually carried out under the micro-canonical ensemble (where the number of particles N, the volume V and the energy E of the system are kept constant) also known as the constant NVE ensemble. Biological systems (as others) are often tremendously affected by both the temperature and pressure of the system and it is usually advantageous to

use a constant temperature (NVT) or constant pressure and temperature (NPT) scheme to produce realistic dynamics.

By controlling the temperature or pressure to produce an NVT or NPT scheme, care must be taken when choosing a thermostat which will maintain the simulation at a desired temperature (often 300 K, but it depends on the process of interest) without causing too much disruption to the dynamics of the system or constraining it in an unrealistic manner and this can be difficult. Of particular interest in this work is the Langevin scheme which adds a random force $\mathbf{W}_i$ and a frictional force with coefficient ζ to the particles in the system at a rate determined by the collision frequency γ_i. These are related by $\gamma_i = \frac{\zeta}{m_i}$ where m_i is the mass of particle i. This scheme simulates random collisions between the system and an external heat bath which will be at the desired temperature. The equation of motion then becomes:

$$m\ddot{\mathbf{r}}_i = \mathbf{F}_i - \gamma_i \mathbf{p}_i + \mathbf{W}_i \tag{2.31}$$

Here $\ddot{\mathbf{r}}_i$ is the acceleration of particle i, $\mathbf{W}_i$ is uncorrelated in time and has zero mean and a variance given by:

$$\sigma^2_{\mathbf{W}} = 2m_i\gamma_i K_b T \tag{2.32}$$

where K_b is the Boltzmann constant and T the desired temperature. The advantage of this scheme is that it allows for stable dynamics while still reproducing the correct canonical ensemble unlike schemes such as the Berendsen thermostat which uses velocity scaling instead to re-scale the velocities of all the atoms in the system to fit to the desired temperature.[97] A drawback of the Langevin thermostat as with other stochastic temperature schemes is that momentum transfer is modified and so it is not possible to compute properties such as diffusion coefficients.[98, 99]

Similar to thermostats, there are a variety of barostats available depending on requirements of the simulation and care must be taken here as well. The Berendsen barostat is commonly used in order to maintain constant pressure in isotropic systems (such as for a solute dissolved in water) and uses a scale factor μ which is a function of pressure to suitably rescale the

system.[97]

$$\mu = \left[1 - \frac{\Delta t}{\tau_P}(P - P_0)\right]^{\frac{1}{3}} \tag{2.33}$$

In Equation 2.33 Δt is the time-step used, τ_P is a time constant of the barostat which determines how closely coupled the system will be to the desired pressure P_0.[97]

A number of force fields have been developed for a variety of systems over time. As these are usually produced at least partially from empirical data, they perform better for some systems over others. In this work we are most interested by those developed for biological study, primarily those based off the original Cornell et al. ff94 force field[100] (which in itself was a development from earlier efforts by Weiner et al.[101, 102]). These force fields fit charge parameters from Restrained Electrostatic Potential (RESP) calculations, Van der Waals parameters calculated from liquid simulations and equilibrium bond parameters from experimental sources (microwave, neutron diffraction and NMR studies for example) and attempt to accurately describe nucleic acids. Over time further developments have led to the creation of the ff99SB force field[103] which attempts to improve secondary structures and backbone dihedrals of glycine and alanine. A further development of this force field was ff99SB-ildn[104] where the side-chain torsion of isoleucine, leucine, aspartate and asparagine are further refined to better describe α-helical and other secondary structures. An alternative development which we also investigate is the Duan et al. 2003 force field[105] (ff03) which varies from the ff99 case in that charges are derived using a continuum dielectric to mimic solvent effects. The backbone protein torsions ϕ and ψ have also been modified. Later, ff03r1 was constructed with enhanced charge parameters for terminal chain protein residues which had previously been taken from the ff94 field. The last force field we investigate is ff12SB which modifies the ff99SB-ildn force fields to include side-chain torsion corrections in an attempt to compensate for this force fields tendency to under-stabilise helical conformations of transiently folded peptides. This force field therefore includes corrections to the ϕ' and ψ' torsions for nucleic acids on top of the other corrections included in ff99SB-ildn.[95]

2.4 Free Energy Exploration and Estimations

2.4.1 Metadynamics

Many biological effects of interest (such as protein folding) only occur on relatively large time scales ($10\mu s$ - $1ms$ or more), whereas current limits on accessible time scales is on the order of 1 ms [106] though this was only possible on a custom built machine. This makes it effectively impossible to regularly reproduce these kinds of processes on atomistic scales. Even when implementing simplifying methods such as SHAKE and Ewald summations, the total CPU time required for a moderately sized molecule to fully explore a region of phase space can be extremely large. In many cases it is not possible or convenient to simulate a system long enough for a desired process to take place and so many attempts have been made to accelerate these 'rare events'. One such method, which has proved successful since it was proposed by Laio and Parrinello in 2002, is Metadynamics [107] which builds up a history dependent potential alongside a standard MD simulation. This potential is based upon a set of a few collective variables (CVs) in order to force the system out of a meta-stable state (some local energy minima) enabling evaluation of free energy surface (FES) of the system with respect to these CVs. These CVs may be an inter-atomic bond length, a torsional or bond angle or other property of the system but it is crucial that they give a good, general description of the process of interest and evolve slowly in time: so great care must be taken with their choice. The equilibrium behaviour of these variables are completely defined by the probability distribution:

$$P(s) = \frac{exp\left(-\left(1/k_B T\right) F(s)\right)}{\int ds\, exp\left(-\left(1/k_B T\right) F(s)\right)} \tag{2.34}$$

where s is the d dimensional vector representing the values of the CVs and the free energy $F(s)$ is given by:

$$F(s) = -k_B T\, ln\left(\int dx\, exp\left(-\frac{1}{k_B T} V(x)\right) \delta(s - S(x))\right) \tag{2.35}$$

where $S(x)$ is a function of the coordinates of the system while s is the

value of the CVs. Using these CVs a bias potential is built up of repulsive Gaussians alongside the standard MD (these Gaussians are deposited along the CV trajectory at some appropriate time interval) which eventually forces the system out of its current region of phase space. As many Gaussians are deposited it will eventually become possible for normal thermal fluctuations to break the system out of its local minima and into a new one. It is important that the Gaussians are the proper size to compromise between sampling speed and accuracy of the reconstructed energy surface and so the height ω and width δ_s of these Gaussians is determined by the typical variation of this surface. The history dependent potential is given by

$$V_G(S(x),t) = \omega \sum_{\substack{t'=\tau_G,2\tau_G,... \\ t'<t}} exp\left(-\frac{(S(x)-s(t'))^2}{2\delta_s^2}\right) \tag{2.36}$$

where τ_G is the frequency of the Gaussian deposition and $s(t)$ is the value taken by the CVs at time t. If the Gaussians used are too large, detail is lost about the free energy surface. If, on the other hand, the Gaussians are too small the simulation may never reach the 'rare event' of interest or alternative energy minima. It is important to note that this method is limited by the number of CVs used as the time required to fill the free energy surface is proportional to $(\frac{1}{\delta})^n$ where n is the number of CVs used.[3] For a sufficiently long time it is assumed that

$$\lim_{t\to\infty} V_G(s,t) \sim -F(s) \tag{2.37}$$

This relation was originally postulated heuristically from observations of the effects of V_G on the dynamics of the CVs on known free energy surfaces but has since been proved.[3] Therefore over the course of a metadynamics simulation the entire region of phase space searched is filled up with these Gaussians and the free energy surface of the system can be reconstructed. An example of this is shown in Figure 2.1.

An extension of this method known as Well-Tempered (WT) Metadynamics attempts to address one of the main problems of the method, namely that free energy surface does not converge but fluctuates around the correct value making it difficult to know when the metadynamics run should end

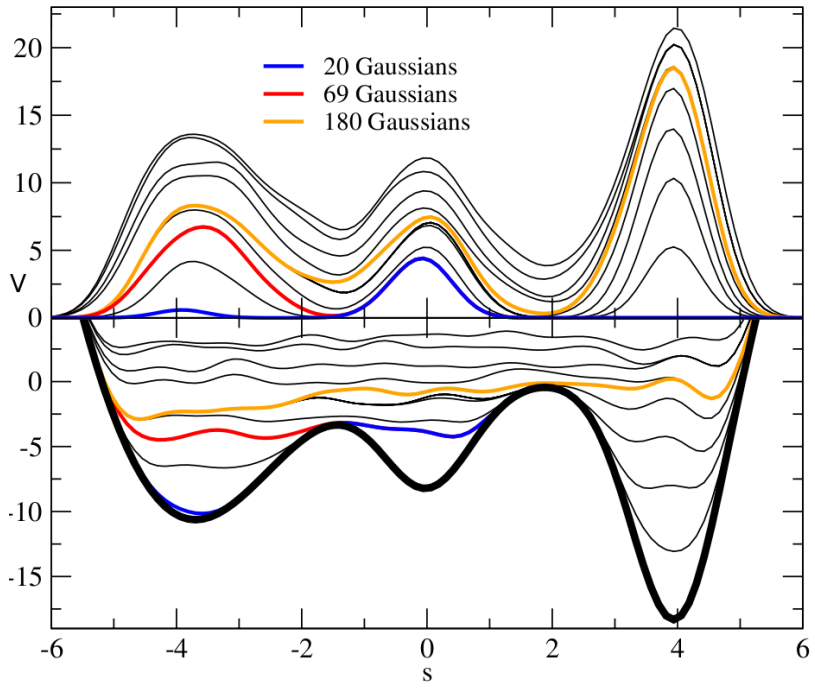

Figure 2.1: An example metadynamics simulation. Evolution of the potential of a 1-dimensional system with a 3-minima potential as shown with respect to the collective variable s in the bottom panel. A biased potential V_G is included and its evolution over time is shown by the coloured lines at different stages of the metadynamics simulation. The simulation begins in the middle minima and after 20 Gaussians (blue line) escapes to the second minima (left). After 69 Gaussians (red) this minima is also filled and eventually the system is able to escape into the true minimum (right). After 180 Gaussians (orange) the potential wells have been filled and the metadynamics becomes diffusive. The top panel shows the energy surface reconstructed using these deposited Gaussians.[3]

as well as impeding the reproduction of an accurate FES. Thus in the WT metadynamics scheme the bias growth rate is changed by adding a history dependent potential:

$$V(s,t) = \Delta T ln\left(1 + \frac{wN(s,t)}{\Delta T}\right) \tag{2.38}$$

where ΔT is a temperature, w has the dimension of an energy rate and represents the initial bias deposition rate and $N(s,t)$ is the histogram of the CV value from an unbiased simulation. The height of each Gaussian is

then determined by:

$$\omega = we^{-\left[\frac{V(s,t)}{\Delta T}\right]\tau_G} \tag{2.39}$$

which over a long period of time allows the metadynamics run to converge to an estimate of the free energy:

$$\lim_{t\to\infty} V(s,t) = -\frac{\Delta T}{T+\Delta T}F(s,t) \tag{2.40}$$

where T is the temperature of the system and $\frac{T+\Delta T}{T}$ is the bias factor to be chosen, which determines the speed at which the bias potential decreases with time.[3, 108, 109]

One possible use of metadynamics in biological systems is the recently developed funnel metadynamics for simulating the long time scales of protein-ligand binding.[110] Here, by using a restraint potential which has a funnel shape and reduces the sampling space of the process of interest to the immediate area around the binding site, it is possible to obtain the details of protein-ligand binding/unbinding including an estimate of the binding free energy as well as the effects of mediating waters.

2.4.2 Poisson-Boltzmann Surface Area Calculations

When considering molecular systems, especially large macromolecules like proteins in complex with ligands, one of the most important attributes which we investigate is the binding free energy. For systems in the NPT ensemble the corresponding energy is the Gibb's free energy of binding ΔG_{bind}. Treatment of these large macromolecules is usually undertaken classically as described in Section 2.3 where we might be investigating some property of protein-ligand binding for example. In this case we would ideally like to calculate the absolute free enegy of binding ΔG^0_{bind} but this is a rare event on the order of milli-seconds and explicit treatment of the solvent environment means that there will be large fluctuations from solvent-solvent interactions from one frame to the next making this calculation extremely diffcult to converge.

Instead, an effective method is to calculate the free energy of solvation. The Molecular Mechanics Poisson-Boltzmann Surface Area (MM-PBSA)

method is widely used and has been tested for a variety of systems, especially biomolecules.[111–117] With this method, we divide up the total ΔG^0_{bind} into components of vaccuo and solvent free energies of the receptor, ligand and complex to find the solvation free energy:

$$\Delta G^0_{bind,solv} = \Delta G^0_{bind,vacuo} + \Delta G^0_{solv,complex} - (\Delta G^0_{solv,ligand} + \Delta G^0_{solv,receptor}) \tag{2.41}$$

where $\Delta G^0_{bind,solv}$ is the solvation free energy of the system, $\Delta G^0_{bind,vacuo}$ is the free energy required to bind the receptor-ligand system in vacuo, $\Delta G^0_{solv,complex}$ is the free energy difference between the bound complex in solvent and vacuo, $\Delta G^0_{solv,ligand}$ and $\Delta G^0_{solv,receptor}$ are the free energy difference between the ligand and receptor going from vacuo to solvent respectively. To evaluate these solvation free energies themselves we commonly divide into electrostatic and hydrophobic portions:

$$\Delta G_{sol} = \Delta G_{elec} + \Delta G_{vdw} + \Delta G_{cav} \tag{2.42}$$

with ΔG_{elec} as the electrostatic component, ΔG_{vdw} is the van der Waals interaction between solvent and solute (which can be further divided into repulsive and attractive components) and ΔG_{cav} is the free energy required to form the solute cavity within the bulk solvent. This last term incorporates the entropic penalty from the reorganisation of solvent molecules around the solute as well as the pressure exerted by the solute against the solvent to create the cavity. The vacuo free energy, $\Delta G^0_{bind,vacuo}$, is obtained via calculating the average interaction energy between the receptor and ligand as well as the entropy change upon binding.

$$\Delta G^0_{vacuo} = \Delta E^0_{mm} - T\Delta S^0_{nm} \tag{2.43}$$

Here ΔE^0_{mm} is the interaction energy calculated from the molecular mechanics force field and ΔS^0_{nm} is the entropy of the system which includes the change of translational and rotational freedom of when moving from one state to the other. The entropy is not always included; it is usually evaluated by normal mode analysis which can be extremely time consum-

ing and for similar shaped and sized ligands, produces very similar results. Instead, for nearly identical ligands, the enthalpy alone can be compared to provide a relative ranking of binding affinity. The interaction energies are calculated from a set of uncorrelated snapshots of an MD trajectory that has already been equilibrated.

We now discuss obtaining the electrostatic contribution to the solvation free energy, ΔG_{elec}.

2.4.2.1 Poisson-Boltzmann Equations

A popular methodology for calculating the electrostatic component of the solvation free energy is to use the Poisson-Boltzmann equations. In this regime, the solute material (our macromolecule for example) is treated as a single body of low dielectric, while the solvent surrounding is considered as a continuum of high dielectric. By defining a system in this way the Poisson equation may be used to give the variation in potential ϕ which in reduced electrostatic units (eliminating the factor $4\pi\varepsilon_0$) is:

$$\nabla^2\phi(\mathbf{r}) = -\frac{4\pi\rho(\mathbf{r})}{\varepsilon} \tag{2.44}$$

where ε is the uniform dielectric constant and ρ is the charge density. In the case of a set of point charges in a constant dielectric this equation reduces to Coulomb's law (as the charge density is given in C m^{-3}). If this is not the case and the dielectric varies with position (such as with our previously defined solute/solvent case) then the Poisson equation is instead:

$$\nabla \cdot \varepsilon(\mathbf{r})\nabla\phi(\mathbf{r}) = -4\pi\rho(\mathbf{r}) \tag{2.45}$$

The total charge density includes the solute charge density (which classically is contained entirely within the solute cavity) and the ionic charge density externally, thus $\rho = \rho_{int} + \rho_{ext}$. The external dielectric is commonly set to 80 for water, while the solute value varies from 2-4, though 1 is also often used.[111, 118, 119] In a dynamical situation there are usually mobile ions present, such as the small salt concentration in biological environments or counter-ions in simulation, which must be incorporated into this equation. Thermal motion as well as the repulsion of like charges ensure that

the ions will not simply sit at locations of largest electostatic potential and thus we describe the position of these ions with the Boltzmann distribution:

$$n(\mathbf{r}) = \mathcal{N} exp\left(\frac{-W_i(\mathbf{r})}{k_B T}\right) \tag{2.46}$$

where $n(\mathbf{r})$ is the number density of ions at location $\mathbf{r}$, $\mathcal{N}$ is the bulk number density and $W_i(\mathbf{r})$ is the work required to bring the ion, i, from infinity to position $\mathbf{r}$. Given that we can have both positive and negative mobile ions, such as Na^+ and Cl^- for a standard salt solution, we divide Equation 2.46 into the positive and negative distributions of ions of charge $\pm e_c$, $W_i = +e_c\phi(\mathbf{r})$ and $W_i = -e_c\phi(\mathbf{r})$:

$$n^+(\mathbf{r}) = \mathcal{N} exp\left(\frac{-e_c\phi(\mathbf{r})}{k_B T}\right), n^-(\mathbf{r}) = \mathcal{N} exp\left(\frac{e_c\phi(\mathbf{r})}{k_B T}\right) \tag{2.47}$$

which represent the distributions of positive and negative ions respectively. We have assumed here that the ratio of ion concentration is approximately 1:1 far from the solute. Therefore, by combining Equations 2.45 and 2.47 we get the non-linear Poisson-Boltzmann equation:

$$\nabla \cdot \varepsilon(\mathbf{r})\nabla\phi(\mathbf{r}) - \kappa' sinh[\phi(\mathbf{r})] = -4\pi\rho(\mathbf{r}) \tag{2.48}$$

where we have incorporated most of the terms into κ', related to the Debye-Hückel inverse length, κ, by $\kappa^2 = \frac{\kappa'^2}{\varepsilon} = \frac{8\pi N_A e^2 I}{1000\varepsilon k_B T}$ where e is the electronic charge, I is ionic strength of the solution and N_A is Avogadro's number. At low ionic concentrations (as in most biological cases) it is acceptable to linearise Equation 2.48 by writing the hyperbolic sine function as a Taylor expansion and keeping only the first expansion, giving:

$$\nabla \cdot \varepsilon(\mathbf{r})\nabla\phi(\mathbf{r}) - \kappa'\phi(\mathbf{r}) = -4\pi\rho(\mathbf{r}) \tag{2.49}$$

In order to solve Equation 2.49 several approaches have been developed of which one of the most favoured numerical approaches is the finite difference Poisson-Boltzmann method.[120–123] In this case, cubic grid lattices are superimposed over the solute and solvent system and the electrostatic potential, charge density, dielectric constant and ionic strength are calculated at each grid point. Although it is possible to assign a grid for atomic

resolution (or higher) charge allocation is usually spread over the six or eight nearby points in a distance dependent manner (as ions will rarely exist exactly on a grid point). The potential at a grid point, k, is then:

$$\phi_k = \frac{\sum_i \varepsilon_i \phi_i + 4\pi \frac{q_k}{L}}{\sum_i \varepsilon_i + f(\phi_k)\kappa_0'^2 L^2} \tag{2.50}$$

where L is the side of the grid cube, ϕ_i is the potential at the nearest neighbours i, ε_i is the dielectric constant midway between ϕ_k and ϕ_i, $f(\phi_k)$ is $sinh[\phi(\mathbf{r})]$ for the non-linear PB equation and 1 for the linear case. $\kappa_0'^2$ is zero inside the solute cavity and κ'^2 elsewhere.

This method requires careful assignment of the solute-solvent boundary, most often using either the molecular surface or the accessible surface, in order to correctly give solute and solvent dielectric values to the grids. Once assigned it is possible to give each grid point its associated internal or external dielectric constant. In this way it becomes possible to calculate a value for the ΔG_{elec} in Equation 2.42:

$$\Delta G_{elec} = \frac{1}{2} \sum_i q_i (\phi_i^{80} - \phi_i^1) \tag{2.51}$$

where two seperate calculations are performed, one in vacuo (solvent dielectric 1) and one in water (solvent dielectric 80), and the electrostatic energy of a charge q_i in a potential ϕ_i is equal to $q_i\phi_i$.[66]

The boundary of the finite grid must be given special care and most implementations of the finite-difference method will also include a form of focussing, whereby the calculation is done with a large continuum solvent with the solute far from the grid boundary to obtain a coarse potential map. A focussed calculation is then undertaken where the solute occupies a larger percentage of the grid volume, so that all of the boundary points of the new grid were internal grid points of the more coarse grid. The new grid points then use the coarse grid values as initial guesses which should give a much better treatment of the boundary.

While the finite difference method has been successful for many systems, it does have some drawbacks. The treatment of the solute-solvent interface for example can be a problem as in many cases one cannot really consider the water interface as a mobile bulk liquid. It can also be diffi-

cult to capture the complex electrostatic effects of a solute macromolecule which are ususally anisotropic and include several regions of high polarity as well as hydrophobic characteristics. One must also take great care with the definition of the solute-solvent interface.

2.4.2.2 Solvent Accessible Surface Area

The solvent accessible surface is usually preferred to delineate the solute-solvent boundary, using a water probe (commonly of radius 1.4 Å) with the values for atom and charge radii taken from some reasonable force field (see Subsection 2.3.1). This adequately describes the electrostatic interactions of the system however inclusion of the non-polar contribution to free energy is also required which arises from short-range interactions between the solute and solvent as shown in Equation 2.42. Solvent exposed area models are also often based on the accessible surface area of the macromolecule and are collectively known as Solvent Accessible Surface Area (SASA) models.[124, 125]

These attempt to calculate the values ΔG_{vdw} and ΔG_{cav} which we saw from Equation 2.42 and are usually given as:

$$\Delta G_{cav} + \Delta G_{vdw} = \gamma A + b \tag{2.52}$$

where A is the total solvent accessible area and γ and b are constants based upon empirical results (usually the experimental free energies obtained when transferring alkanes from vacuo to solvent).

2.5 Protein - Ligand Docking

Many biological processes involve the binding of a molecule to the binding site of a target protein in order to produce some desired effect. This can include processes such as muscle contraction, signalling systems between cells or possibly inhibiting some competing interaction. Most drugs affect the body in a manner similar to this. It is thus of critical importance to enhance drug discovery to be able to locate and hopefully predict these binding sites. Docking programs are one method to attempt to solve this problem:

they can be used to refine a known 3D structure from experiments, or they can be used to perform 'blind' dockings where the binding site is not known. Docking simulations comprise of a form of search function, which searches for the most energetically favourable pose of the ligand around the protein, and scoring system which can be used to rank the placements for the bound system. Usually the scoring function is related to the affinity for binding of the system as this is likely to produce the most stable structure.[126] A common type of scoring function will be similar in form to that of the force field energy equation seen in Equation 2.27.

In this work the Autodock program has been used.[127, 128] This creates a set of grids around the receptor molecule by calculating the energy of each of the ligand atom types at each grid point. Using this the docking algorithm then attempts to find the global minimum of the system by positioning the ligand on the grid. To start, a random position is usually chosen and the ligand is moved around the conformational space of the system. Moves are accepted on criteria specified by the user and are usually based on the ligand having a more favourable potential energy than in the previous step, if this is not true the move is rejected and the docking program tries another move. Eventually the ligand reaches the maximum number of steps specified or alternatively is unable to successfully move into another position and the docking run is completed. Lamarckian genetic algorithms are commonly used and have been shown to produce high efficiency and reliability.[129] These are similar to the more standard genetic algorithm except that the parent conformer, instead of simply transmitting random mutations to its offspring, will pass on some of its own information allowing the conformer to continually refine its own position around the receptor.

There may be many docking runs per docking simulation in order to fully explore the conformational space of the system and the final structures can be clustered via the square root of the mean squared displacement (RMSD) calculation which groups similar conformers together. Often the lowest energy structure of these clusters is then taken to be the 'best' structure in the cluster and several of these can be compared to the particular criteria of the user and used in further computational work. If nothing is known *a priori* about the system then the lowest energy cluster (or possibly a struc-

ture from the largest cluster, where these are not the same) is commonly used.[127]

Due to the computational expense of these codes a common simplifying measure is to consider the receptor entirely rigid (as is done in blind dockings) which reduces the degrees of freedom of the system and makes the calculations more manageable but greatly reduces accuracy. Some or all of the bonds on the ligand are set to be rotatable which allows the ligand to change shape to fit around areas of the receptor and enables a closer docking than would be possible with an all rigid structure. It is also possible to designate flexible receptor residues in a similar manner to the ligand which should further improve the final docking. This significantly increases simulation time due to the extra degrees of freedom in the system and so this is used only when information on the binding site can be attained from previous experimental work (such as X-ray crystallography or NOESY data).

Chapter 3

Structural Properties of Green Tea Catechins

Chapter Overview

In this chapter the structural properties of four green tea catechins are investigated. A suitable force field is parametrised and used for classical molecular dynamics simulations for the catechins in the gas phase and in solution. The conformational free energy landscape of the catechins is explored with metadynamics. Since we are interested in studying the interaction of green tea catechins with the cardiac protein troponin C (Chapters 5 and 6), we also analyse the calcium sensitiser EMD 57033 and Levosimendan through similar techniques.

3.1 Green Tea Catechins

As has been discussed in detail in Chapter 2, in addition to the classically attributed calmative benefits of tea, much speculation and research have gone into investigating its possible role in preventing or ameliorating the effects of chronic diseases, especially its possible preventative effects against heart disease, diabetes, Alzheimer's, Parkinson's and cancer.[8, 25, 26, 130, 131] There are a wide variety of mechanisms of action suggested to account for these effects, notably including the antioxidant and pro-oxidant effects of the green tea polyphenol catechins present in green tea. Previous studies have

established that of the many available catechins present in green tea, the majority of the active polyphenol content lies in only a few catechins, four of which were chosen for analysis: (—)-Epigallocatechin-3-gallate (EGCg), (—)-Epicatechin-3-gallate (ECg), (—)-Epigallocatechin-3-O-(3-O-methyl)-gallate (EGCmg) and (—)-Epigallocatechin (EGC). They are shown in Figure 3.1 with labelled benzopyran core ring (A/C) as well as gallate with 3,4,5-trihydroxy groups (B) and galloyl moiety (B′) with ester form.[131] The dihedral torsions which determine the conformational layout of these catechins are also labelled γ, θ and ϕ. These are of interest as they determine the respective positions in or out of plane with respect to one another of the aromatic rings which themselves contain several hydroxyl groups prone to interact with the local environment via hydrogen bonding. The rings are capable of cation-π or $\pi - \pi$ stacking interactions which may also be important in biological environments such as in binding to proteins.

In order to fully understand the interactions of these molecules in solvent and protein environments, classical simulations are required as the system sizes and time-scales involved make more accurate *ab initio* methods prohibitively expensive. The remainder of this Chapter is dedicated to the parametrisation of a suitable classical force field and interrogation of the structural aspect of these molecules from molecular dynamics and metadynamics simulations.

3.2 Force Field Parametrisation for Classical Simulation

Initially, total energy DFT calculations of the selected green tea catechins were undertaken using Gaussian09[132]. The three dihedral angles deemed to determine the interactions available to the catechin molecules (as shown in Figure 3.1) were analysed using the B3LYP exchange and correlation functional with a 6-311++G(d,p) basis set. This was in order to give suitable parameters to compare with the derived classical model. Potential energy profiles were calculated whereby these angles were scanned at 22.5° increments with the remaining two angles restrained to minima positions using harmonic potentials. In several minima positions single point cal-

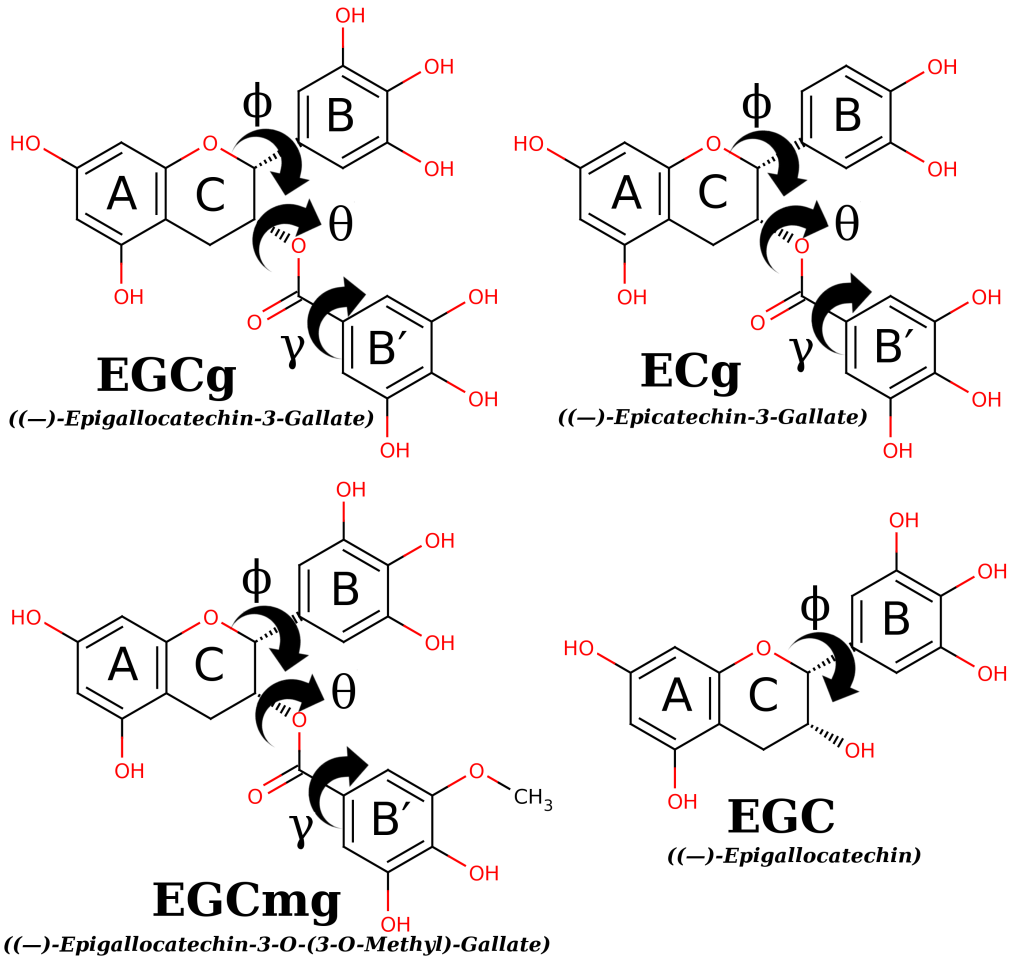

Figure 3.1: The four catechins selected. Included are labelled rings and torsional angles of interest.

culations using the Møller Plesset method truncated to the second order (MP2)[70, 133–137] were carried out to determine highly accurate torsional energy values for the γ torsion at the maxima (-90° and 90°) and minima (0° and 180°) values in order to confirm the relative barrier height. Hartree-Fock single point energies for the B3LYP optimised structures were also calculated in order to produce a parametrisation which would be compatible with GAFF (General AMBER force field).[138] RESP (Restrained Electrostatic Potential) charges were finally used as the classical point charges of the molecules. RESP charge derivation was carried out in AMBER[139] using ESP charges from eight optimised minima positions which corresponded to

the possible meta-stable states of the dihedral torsions derived from the potential energy scans (or 2 for EGC which lacks the γ and θ dihedral angles). In order to determine whether this classical force field described the system well, relevent torsion angles were again scanned at 22.5° increments in AMBER in order to determine whether the DFT torsional profile were reproduced. Harmonic potentials were used to constrain two angles at previously established energy minima (for $\gamma = 0°$, $\theta = 112.5°$ and $\phi = 90°$) while the other angle was rotated at 22.5° increments as in the DFT calculations. The molecule was minimised with the conjugant gradient method at each angle scanned. Using the default GAFF parameters for the molecular torsions in AMBER led to the correct minima positions but overestimation of the γ torsional barrier by a factor of 4. This was found to be due to the lack of a specific dihedral parameter in GAFF for atom types Os-C-Ca-Ca which was automatically replaced by a default X-C-Ca-X value derived from C_6H_6. In order to correct this, a new dihedral parameter was generated to more accurately describe this torsion with $IDIVF = 1$, $PK = 1.1$ (kcal/mol), $phase = 180°$ and $n = 2$ where the functional form of torsion parameters in AMBER is defined as:

$$V_{tors} = \frac{PK}{IDIVF}(1 + cos(n\phi - phase)) \tag{3.1}$$

With the force field correction the γ minima points are well reproduced, while the maxima are within 1.5 kcal/mol of the MP2 value.

The torsional profiles of EGCg, including those obtained with the corrected torsional parameter of γ, can be seen in Figures 3.2, 3.3 and 3.4. When energy minimisation is carried out on these profiles in AMBER one can see some difference in absolute barrier height though the overall shapes of the profiles are preserved. The ϕ torsion shows a sharper peak at 22.5° and $-157.5°$ which is nearly 2 kcal/mol higher than the HF value. This is due to the forced steric overlap of atoms on the B and B′ rings which the AMBER minimiser is unable to completely account for, however during a molecular dynamics simulation without restraints this is not a problem as this conformation will not be encountered. The θ torsion minima at $-67.5°$ is shown to be slightly too large compared to the DFT and HF scan which

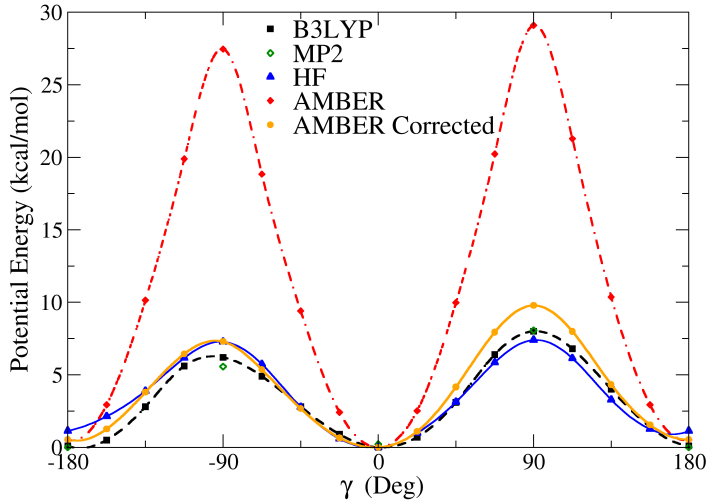

Figure 3.2: Potential Energy profiles for EGCg as a function of the γ angle for $\theta = 112.5°$ and $\phi = 90°$ calculated with DFT-B3LYP and the classical AMBER force field. The B3LYP profile is in black, the HF in blue, the original profile without torsional correction is shown in red while the corrected profile is shown in orange. The Single point MP2 energies at selected angles are shown in green. The zero of the energy scale is set to the absolute minimum.

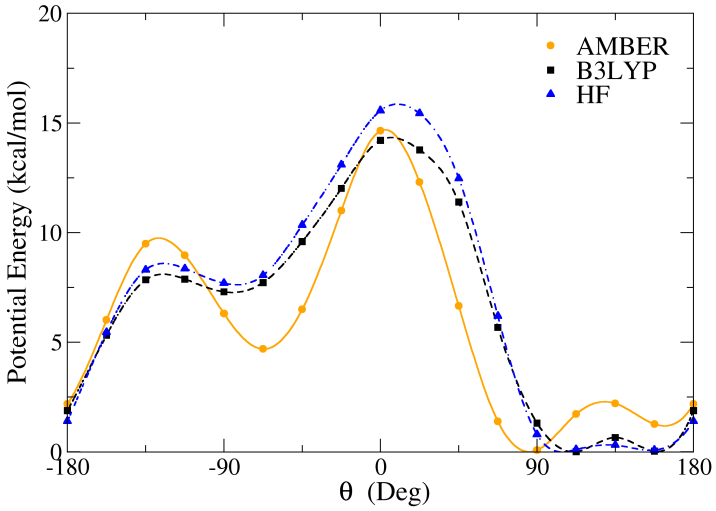

Figure 3.3: Potential Energy profiles for EGCg as a function of the θ angle for $\gamma = 0°$ and $\phi = 90°$ calculated with DFT-B3LYP and the classical AMBER force field. The B3LYP profile is in black, the HF in blue and the classical profile is in orange. The zero of the energy scale is set to the absolute minimum.

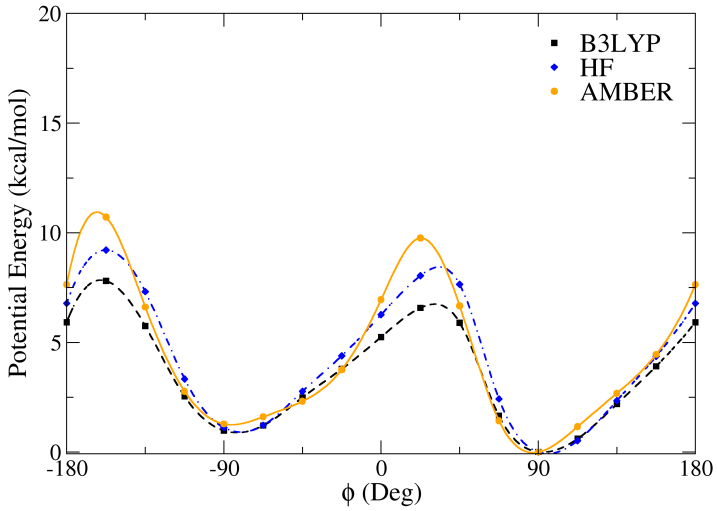

Figure 3.4: Potential Energy profiles for EGCg as a function of the ϕ angle for $\gamma = 0°$ and $\theta = 112.5°$ calculated with DFT-B3LYP and the classical AMBER force field. The B3LYP profile is in black, the HF in blue and the classical profile is in orange. The zero of the energy scale is set to the absolute minimum.

appears to be due to a slightly less disfavoured interaction between the ester oxygen on ring B′ and the nearby benzopyran A ring carbons. The barriers at $-135°$ and $0°$ are reasonably well re-produced (to almost 1 kcal/mol accuracy) however, and indicate that this minima conformation cannot be explored under normal conditions. All the catechin profiles show similar characteristics with the exception of EGC which lacks both the γ and θ angles, leaving the B ring able to freely rotate. Overall, despite these small energy differences in absolute barrier height, the classical force field appears to reproduce the main characteristics of these torsions well and was thus used in the classical simulations.

As discussed in Chapter 2, in some systems dispersion effects can cause DFT functionals to perform badly. In order to test the difference between a dispersion corrected and our uncorrected B3LYP functional, geometry optimisations were performed in Gaussian09 for the EGCg γ torsion at stationary points (-180°, -90°, 0°, 90° and 180°) using B3LYP-D with an S_6 scaling factor of 1.05 according to Grimme's dispersion correction.[86, 87] The γ torsion was chosen as it has well defined, similar barriers in all the

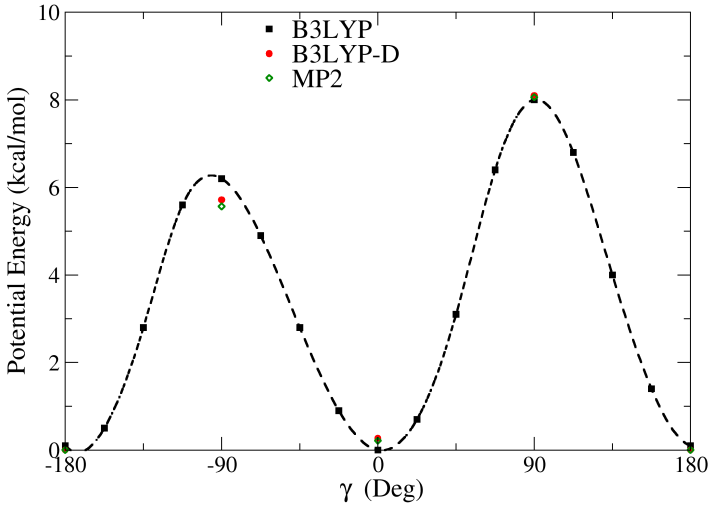

Figure 3.5: Comparison of the γ torsion potential energy profile of EGCg using B3LYP (black), B3LYP-D (red), and MP2 single point calculations for reference (green).

catechins tested and its position affects the proximity between the B and B′ rings. The results are shown in red in Figure 3.5. The main differences are in the 0° minima, where the B3LYP-D optimisation gives an energy 0.26 kcal/mol higher than the original B3LYP, and at the maxima at -90° where the barrier height is reduced by 0.5 kcal/mol. This is also closer to the MP2 single point result. The difference is likely to be due to the fact that the hydroxyl groups on the B′ ring are oriented towards the B ring in this position and a small interaction occurs between the attached hydrogens on the rings. Given the relatively small improvement shown here it appears unlikely that the B3LYP-D functional would change the parametrisation used for the force field and appears reasonable to use the uncorrected value though future work on similar structures may find the improvement valuable.

3.3 Classical Molecular Dynamics of Catechins

3.3.1 Catechins in Vacuo

The initial structures of the catechins were taken from the minimised positions of the profiles discussed in Section 3.2. The major dihedral angles were set to $\gamma = 0°$, $\theta = 112.5°$ and $\phi = 90°$ where appropriate. Vacuo simulations of the catechins were carried out with the AMBER10 package using the parametrisation outlined in Subsection 3.2. Each catechin was initially minimised for 10,000 steps using the steepest descent method for the first 5,000 steps followed by conjugate gradient minimisation for the remainder with a 10 Å cutoff for non bonded interactions. Molecular Dynamics was then used to heat the molecule up to 300 K with a Langevin thermostat using a 1 ps^{-1} collision frequency. With SHAKE[92] switched on to constrain the bonds containing hydrogen atoms, a time-step of 2 fs was used and each simulation run for 50 ns. While SHAKE is not strictly necessary for these calculations as the total number of atoms in the system in each case is less than 100, it allows us to easily compare to later simulations which include water and a larger protein environment. This allowed averages for dihedral motion and intra-molecular hydrogen bonds to be ascertained. Results are displayed in Subsection 3.3.2 where they are compared directly to the solvent simulations. The total intra-molecular hydrogen bonds are listed in Table 3.1 where they are also compared with the water and chloroform solvent cases.

3.3.2 Catechins in TIP3P Water and Chloroform

The catechins were prepared for solvent simulations by surrounding them in a 10 Å buffer of TIP3P model waters[140] periodically repeated in a truncated octahedral box. In this case the minimisation and MD was split into solvent only and full system equilibration phases. First, with the catechin held fixed by positional harmonic restraints, the solvent was minimised using the combination of minimisation steps in Subsection 3.3.1. Following this the restraints were removed and a further minimisation of the entire system was carried out for 10,000 steps as previously stated. The equilibra-

Table 3.1: Average number of Intra- and Inter-molecular hydrogen bonds during the catechin MD simulations. In brackets are the averages obtained when lowering the angle cut-off to 100°.

	Intra			Inter
Catechin	Vacuo	Chloroform	Water	Water
EGCg	0.5 (3.8)	0.4 (3.9)	0.3 (3.7)	16.9
ECg	0.7 (2.9)	0.8 (3.0)	0.3 (2.8)	16.5
EGCmg	0.9 (3.9)	0.9 (3.9)	0.5 (3.6)	16.5
EGC	0.3 (2.5)	0.3 (2.5)	0.2 (2.1)	12.9

tion phase followed this pattern with 20 ps of MD heating the system from 0 K to 300 K using the Langevin thermostat with a 1 ps^{-1} collision frequency, with harmonic restraints on the catechin and SHAKE[92] switched on, to heat the solvent only. Subsequently the restraints were removed and the system allowed to equilibrate for a further 100 ps with the same thermostat at 300 K and a Berendsen barostat with a target pressure of 1 bar and relaxation time of 1 ps. Production runs of 50 ns were then carried out to produce data to contrast the vacuo simulations. Alternatively, to simulate a hydrophobic medium, the catechins were also placed in 10 Å buffers of chloroform ($CHCl_3$) to monitor the change in intra-molecular bonding and preferential dihedral conformations. The minimisation and equilibration phase was identical to that performed for the TIP3P waters and the simulations were run for 50 ns as previously.

Hydrogen bonding analysis was carried out to determine the behaviour of the catechins in vacuo and solvent. These were tracked throughout the simulations using the ptraj H-bonding facility with the AMBER package. H-bond acceptor heavy elements are specified as well as possible electron donors with both the hydrogen atom and the O or N atom they are connected to. An H-bond is deemed to have been formed if the distance between the donor and acceptor and the O-H-O (for example) angle is within predefined parameters. The inter-atomic distances of all donors and acceptors which are able to form H-bonds were monitored. The donor-acceptor distance cut-off for inter-molecular H-bonds was 3.5 Å and the angle cut-off was set to only accept angles larger than 120°. These results are then aver-

aged throughout the production run simulation. From examination of the minimised positions of the ring hydroxyl groups, as well as their angular fluctuations in vacuo and solvent, it appears that the optimal intra-molecular hydrogen bonding position occurs when the O-H-O angle is 110° and so the H-bonding analysis was also carried out with an angular cut-off of 100° in order to monitor this interaction. These averages are given in brackets in Table 3.1 and are likely to be a better representation of the self interactions occurring in the simulations. This is reasoned to be because inspection of of catechin B and B′ rings would suggest that the hydroxyl groups could form around two hydrogen bonds on each ring, far more than is seen with the 120° cut-off but very close to that of the 100° cut-off.

In Table 3.1 some interesting trends are observed. From vacuo to solvent the intra-molecular H-bonds appear to decrease very slightly except for EGC which is moderate. This is due to the EGC O2 weak H-bond breaking in the solvent case, reducing the total. There is little difference in the overall solvent H-bonding when comparing ECg and EGCmg suggesting that substituent effects on the B and B′ ring will not have much effect to the overall external system interactions. For EGCmg it appears that the methyl group on the attached oxygen strengthens the intra-molecular H-bond to the neighbouring oxygen very slightly compared to EGCg. For ECg there is a significant decrease in the number of H-bonds due to the reduced number of hydroxyl groups on ring B, this may be significant as this was seen to decrease the potency of green tea polyphenols in inhibiting lung cancer growth.[141] In addition it was found that hydroxyls on the galloyl group, ring B′, also showed considerable antiproliferative effects in PC-9 human lung cancer cells[142] suggesting that the additional methyl goup in EGCmg could disrupt this activity. In this way EGCmg and ECg could be considered identical in that they both effectively have one less hydroxyl which can externally interact, leading to reduced activity. A breakdown of H-bonds by catechin oxygen (using the 110° cut-off) is shown in Figure 3.7 while the oxygen numbering system, based upon EGCg, can be seen in Figure 3.6.

Figure 3.7 gives a more in-depth view to the interaction differences between the catechins under different solvent conditions. Vacuo and chlo-

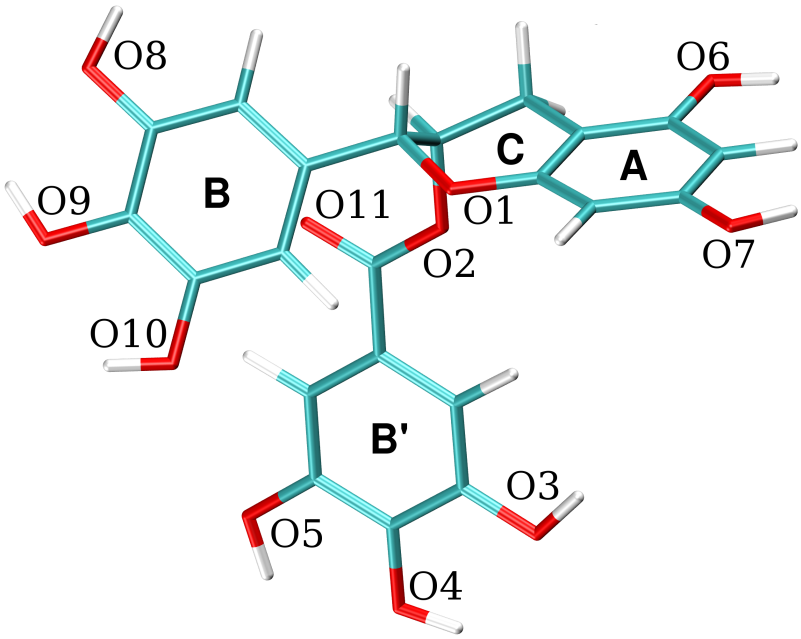

Figure 3.6: EGCg with oxygens labelled according to AMBER nomenclature. In ECg O8 is replaced by a single hydrogen while in EGCmg O3 has a methyl group bound to it instead of the hydrogen. In EGC the B′ ring attached to O2 is replaced with a hydrogen.

roform simulations recover almost identical H-bond breakdowns, primarily through increased hygrogen bonding with the O4 and O9 in the centre of the trihydroxy groups. The presence of water solvent reduces intra-molecular bonding slightly but mainly acts to 'even out' the H-bond intramolecular interaction as in vacuo and chloroform the hydroxyl groups tended not to rotate. In all cases the O6 and O7, present on the A ring, readily form roughly 2.5 bonds each marking them as prime sites for bonding in other systems such as protein or lipid environments. As we might expect, the interactions on the B and B′ rings are very similar with exception of the locations where substitutions on different rings take place. The ECg lack of O8 causes an increased preference for intra-molecular binding to the O9 except in water solvent which shows little difference. Meanwhile the same effect is seen for the EGCmg O4 where the methyl group increases intra-molecular bonding in all environments while drastically reducing it for O3 as we would expect due to the limited potential for rotation.

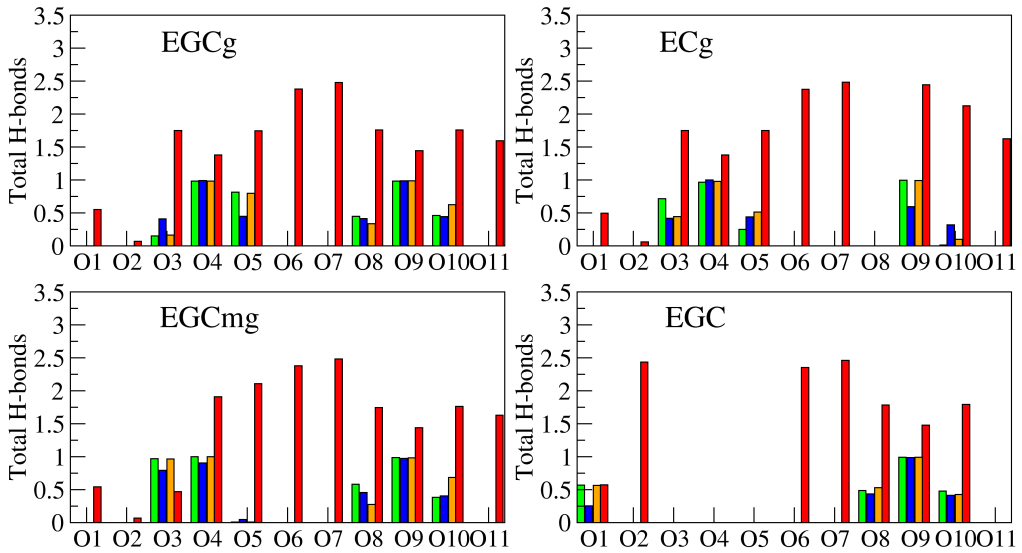

Figure 3.7: H-bonds formed in MD simulation with each catechin oxygen. Green: the number of intra-molecular H-bonds formed in vacuo. Blue: number of intra-molecular H-bonds formed in water. Yellow: number of intra-molecular H-bonds formed while in chloroform. Red: number of inter-molecular H-bonds formed in water. Note that ECg lacks O8, while EGC lacks O3, O4, O5 and O11. All intra-molecular bonds are attributed to their acceptor oxygen to prevent double-counting.

For EGCg, ECg and EGCmg the O1, O2 and O11 oxygens (the ester and benzopyran oxygens) show identical bonding behaviour due to the highly similar configurations they observe. In the case of O1 the low absolute charge on the oxygen as well as the apparent configuration of the B ring co-operatively act to prevent more than a very small amount of water solvation. This is also true for EGC indicating that the B′ ring has little to no effect on this behaviour. While no intra-molecular interactions are seen with O1 in EGCg, ECg and EGCmg in the case of EGC about 0.5 intra-molecular H-bonds occur with O1 from the O2 hydroxyl group present in vacuo and chloroform, reduced to 0.25 in solvent. The lack of the B′ ring also enables significant solvent bonding to O2 in the water simulations though not enough to compensate for the lack of the four other oxygens. Overall we see that water solvation has a strong effect on these molecules and as such, hydrogen bonding is expected to be a major contributor to the interactions

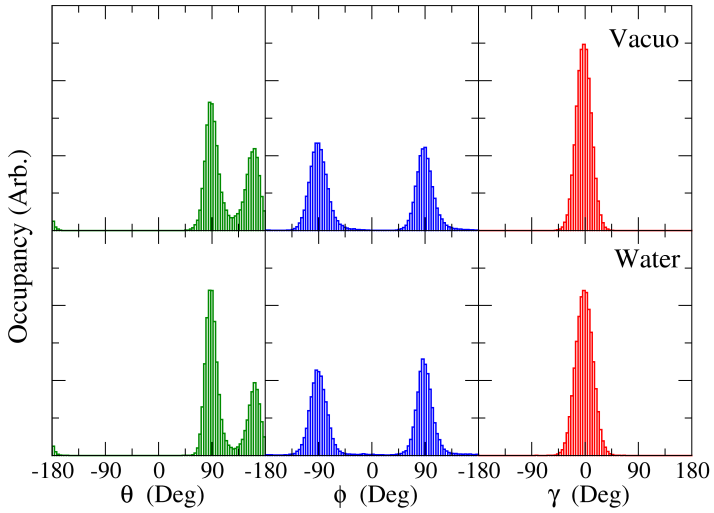

Figure 3.8: Histogram of the major dihedral torsions which determine the structure of EGCg during MD simulation in vacuo (top) and water solution (bottom).

available to these molecules. Due to the number of aromatic rings present in the catechins the possibility of π stacking interactions, especially between the B and B′ rings, was investigated however this was found not to occur due to the conformational restrictions of the rings, primarily due to the B′ ester configuration.

The dihedral torsions which were used to parametrise the system are monitored throughout the 50 ns production runs and shown in histograms in Figure 3.8 for EGCg, Figure 3.9 for ECg, Figure 3.10 for EGCmg and Figure 3.11 for EGC. The effect of the chloroform solvent was seen to be very similar to the vacuo case and so has not been shown.

Overall the dihedral histograms show very similar, if not identical profiles, in accordance with the parametrisation Figures 3.2, 3.3 and 3.4. The γ angle is not seen to switch to the minima at 180° in any simulation while the θ and ϕ angles switch between their two minimas. In all cases the torsions appear to be more restricted in the vacuo case suggesting that H-bonding does indeed play an important role in reducing the torsional barriers. As expected, the lack of the B′ ring on EGC allows much easier rotation and a minimal amount of occupation at all angles is seen, especially in water sol-

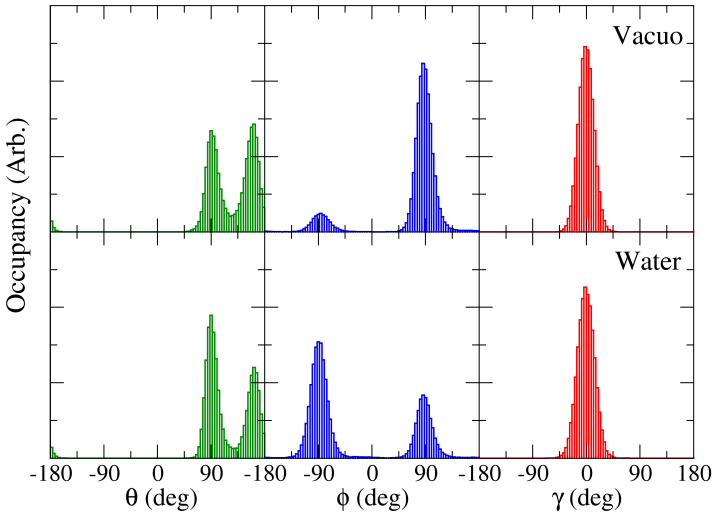

Figure 3.9: Histogram of the major dihedral torsions which determine the structure of ECg during MD simulation in vacuo (top) and water solution (bottom).

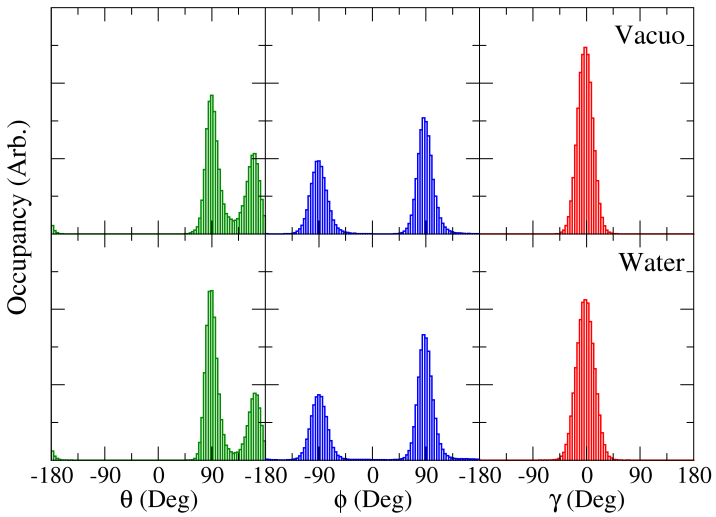

Figure 3.10: Histogram of the major dihedral torsions which determine the structure of EGCmg during MD simulation in vacuo (top) and water solution (bottom).

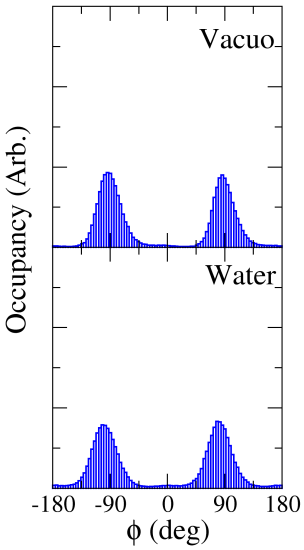

Figure 3.11: Histogram of the ϕ dihedral torsion of EGC during MD simulation in vacuo (top) and water solution (bottom).

vent, though the minima positions at -90° and 90° are still heavily preferred. For the θ angle it appears that the 90° position becomes more preferable in solvent and this is especially noticable in ECg due to the asymmetry of the B ring. In this case it appears that the ϕ and θ angles are heavily influenced by one another with the ester oxygen appearing to slightly interact with the B ring when in the $\theta = 90°$, $\phi = -90°$ configuration.

Radial distribution functions between catechin oxygens and water solvent oxygens have been calculated and can be seen for EGCg, ECg, EGCmg and EGC in Figures 3.12, 3.13, 3.14 and 3.15 respectively. These can be used to give an indication of the change in solvent environment and catechin structure. O6 and O7 are equivalent for each catechin and we see this is reflected in the profiles. As they are far from the other aromatic rings they can interact with water molecules freely (also shown in Figure 3.7 where they show the largest number of H-bonds) and thus reproduce the distinctive first and second solvation shells at 3 Å and 5 Å respectively. O11 also shows identical profiles for EGCg, ECg and EGCmg and shows a clear peak for the first solvation shell while the A and B′ rings exclude further waters, leaving a barely discernable second shell. The tri-hydroxy

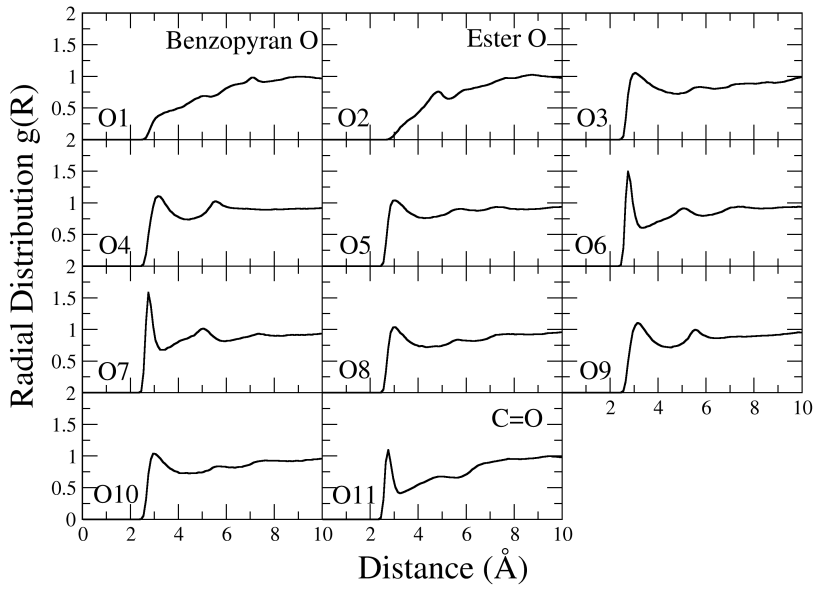

Figure 3.12: Radial Distribution Functions of EGCg oxygens to solvent oxygens.

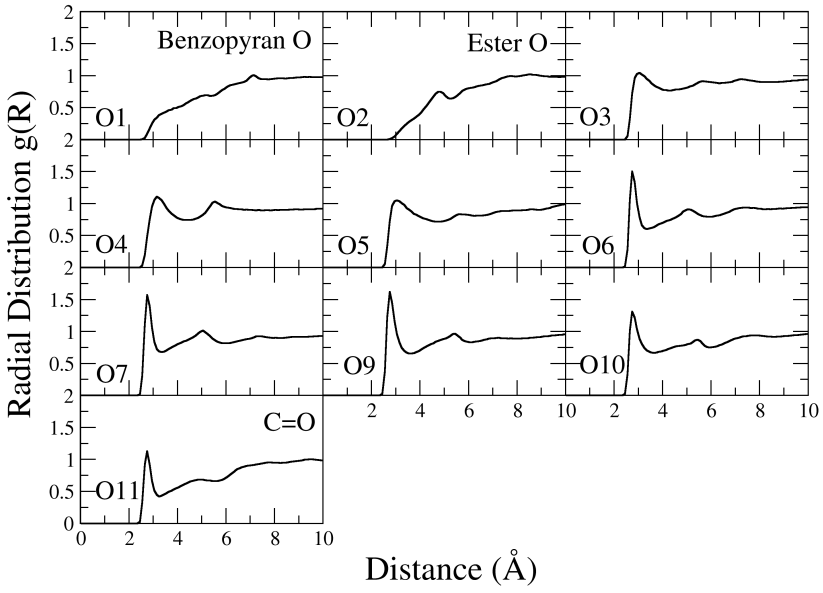

Figure 3.13: Radial Distribution Functions of ECg oxygens to solvent oxygens. The ECg lacks the O8 oxygen which changes the characteristics of the O9 and O10 on the B ring.

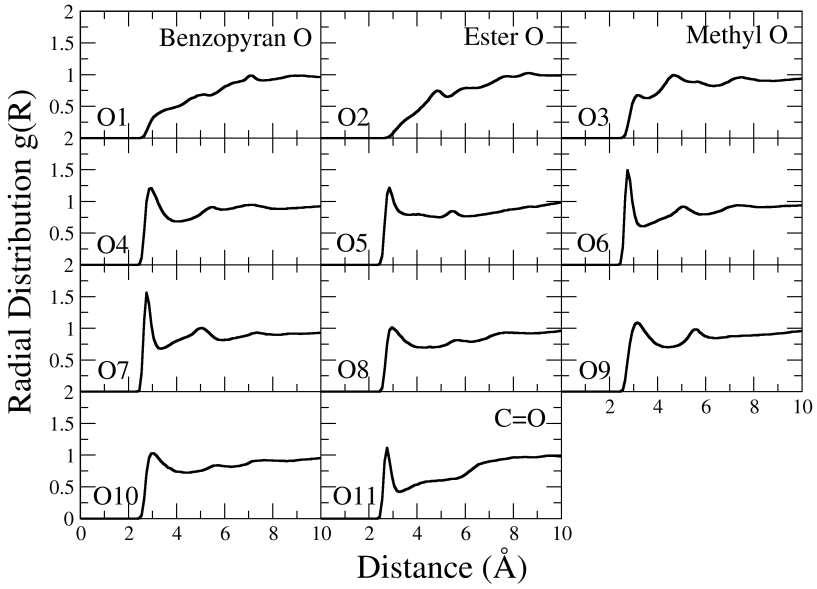

Figure 3.14: Radial Distribution Functions of EGCmg oxygens to solvent oxygens. The EGCmg includes a methyl group attached to O3 which distinctly changes the solvation characteristics of the B′ ring.

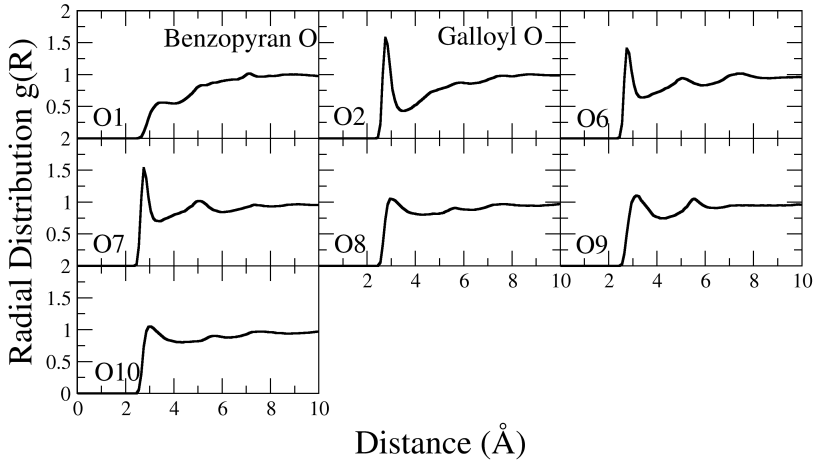

Figure 3.15: Radial Distribution Functions of EGC oxygens to solvent oxygens. Note that the EGCg lacks the B′ ring entirely which is replaced with a hydrogen bonded to O2.

oxygens on the rings (O3, O4 and O5 for ring B′ and O8, O9 and O10 for ring B) show similar characteristics in that the close proximity of the nearby hydroxyls reduces the size of the first solvation peak with respect to O6/7 while the second peak is slightly obscured and at a distance around 5.5 Å. In all catechins the O1 oxygen is seen to be highly isolated due to the nearby rings and thus can only interact minimally with the surrounding waters. This behaviour is also seen for O2 except for EGC where the O2 has become part of a hydroxyl group and is relatively free to interact (and thus the profile is almost identical to the O6 and O7 oxygens). The EGCmg O3 also give a distinct profile due to the attached methyl group and shows that this prevents the close association of waters.

3.3.3 Free Energy Conformational Landscape by Metadynamics

In order to sample the local conformational landscape surrounding the four catechins, metadynamics was undertaken using the previously discussed γ, θ and ϕ torsional angles as collective variables. In the case of EGC only the ϕ angle was selected as the molecule lacks the other two. A Langevin thermostat and Berendsen barostat were used to maintain 300 K temperature and 1 bar pressure using the same settings for the classical MD earlier. Potential energy scans of the torsional profiles in Section 3.2 indicated that the largest energy barriers were on the order of 10 kcal/mol. These were used to determine the choice of Gaussian width and height, which were chosen to be 10° and 0.1 kcal/mol respectively. The initial configuration was chosen to be the equilibrated catechin configurations from the conventional MD. The vacuo and water solvent simulations were chosen for metadynamics. Chloroform simulations were not carried out as the earlier analysis of the MD suggests that vacuo and hydrophobic solvent have effectively identical effects. The Well-tempered scheme was chosen to produce production runs for both vacuo and solvent simulations in order to minimise the error in FES reconstruction using PLUMED.[108] From earlier analysis of the torsional barriers a bias factor of 15 was chosen to overcome barriers of around 8-12 kcal/mol. The FES for solvent and vacuo metadynamics as a function of

CV after 100,000 gaussian depositions can be seen for EGCg in Figure 3.16, for ECg in Figure 3.17 and for EGCmg in Figure 3.18. The free energy profiles as a function of each collective variable integrated over the other two are shown below. The 1-D profiles were produced by integrating out the other two variables using the following general equation:

$$F(x) = -k_B T ln[\int dy \int dz e^{-F(x,y,z)/k_B T}] \tag{3.2}$$

Each reconstructed profile was averaged over the last 1,000 profiles which, when combined with the well-tempering effect reducing gaussian height in flat configurational spaces, acts to further reduce the associated FES reconstruction error to fractions of kcal/mol. The error in these profiles is carefully explored in Section 3.3.3.1. The prominent stationary points of the systems are well reproduced with some slight differences in absolute barrier height. Solvent effects clearly reduce the energy penalty to rotation by as much as 1.5 kcal/mol. This has been determined to be due to the effect of hydrogen bonding on the multiple hydroxyl groups present on the B and B′ rings. The barrier heights of the γ and θ torsions are shown to be far too large (around 8 kcal/mol and 12 kcal/mol respectively) for their rotation to be seen in classical MD. This is corroborated by the dihedral histograms seen in Section 3.2. The effect of substituents on either ring does not appear to greatly impact on the profiles themselves and it appears inter-ring interactions are not changed or enhanced, though there is a very slight interaction of the EGCmg methyl group on ring B′ with ring B which shifts the energy barriers very slightly. The free energy profile of EGC in Figure 3.19 indicates that, except for the EGCmg, the rotation of the B ring is relatively unaffected by the presence or absence of the B′ ring for the catechins as the two rings do not interact with each other. For EGC the presence of solvent shifts the barrier maxima from -30° in vacuo to +30°. This switching effect is not seen with the other catechins and is in any case a small difference of around 0.5 kcal/mol which can easily be overcome by thermal effects. As with the other catechins, these results emphasize the importance of including explicit solvent in the simulations, as the inter-molecular interactions are crucial to the underlying mechanisms which define the catechin environment.

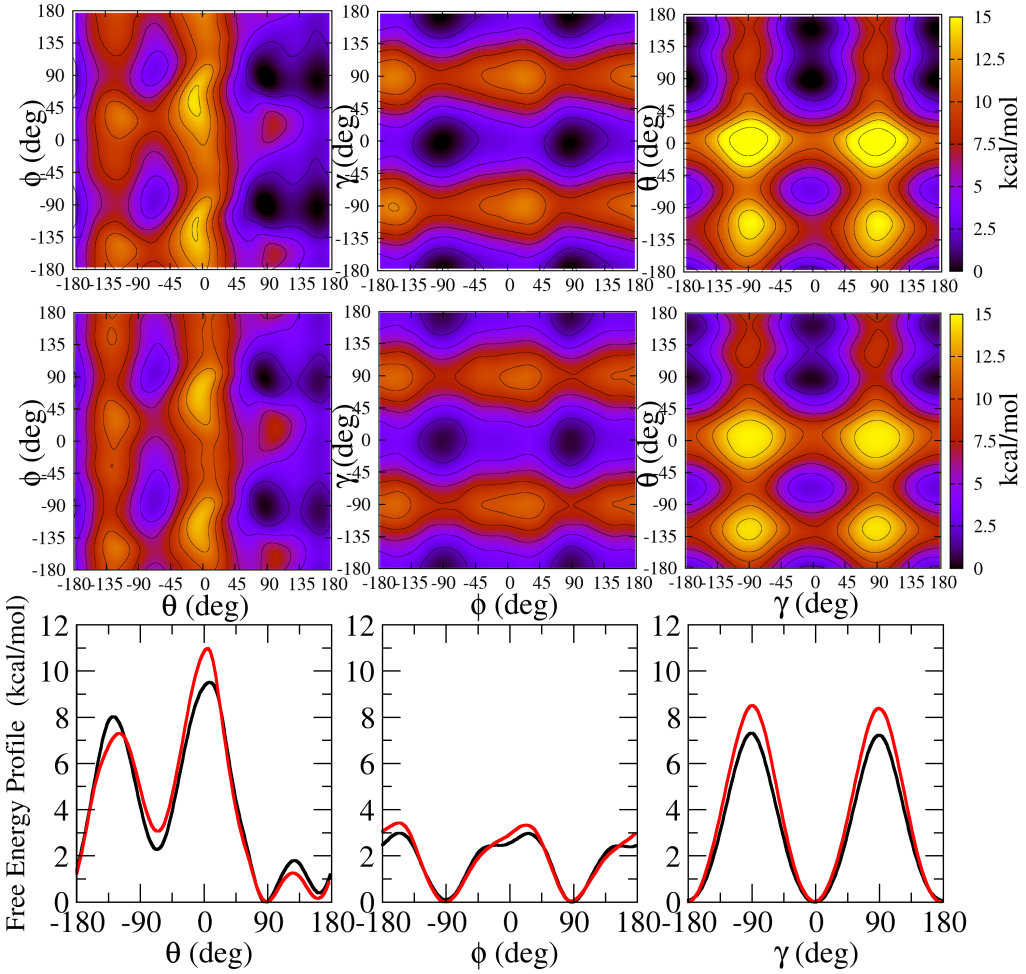

Figure 3.16: Free Energy Surface maps at 300 K as a function of the dihedral collective variables used during metadynamics for EGCg. The top row is the vacuo metadynamics; the middle row is the solvent metadynamics; the bottom row shows the free energy profile as a function of each collective variable integrated over the other two with the black line indicating the solvent profile and the red line indicating the vacuo profile.

3.3.3.1 Analysis of the Error in Metadynamics Profiles

In order to examine the accuracy of the metadynamics profiles produced throughout this work, a series of profiles were produced in a simulation-time dependent manner and then overlaid to show the fluctuational difference as Gaussians are deposited. The profiles were produced for every 10,000 Gaussians deposited and those profiles were averaged over the last 1,000

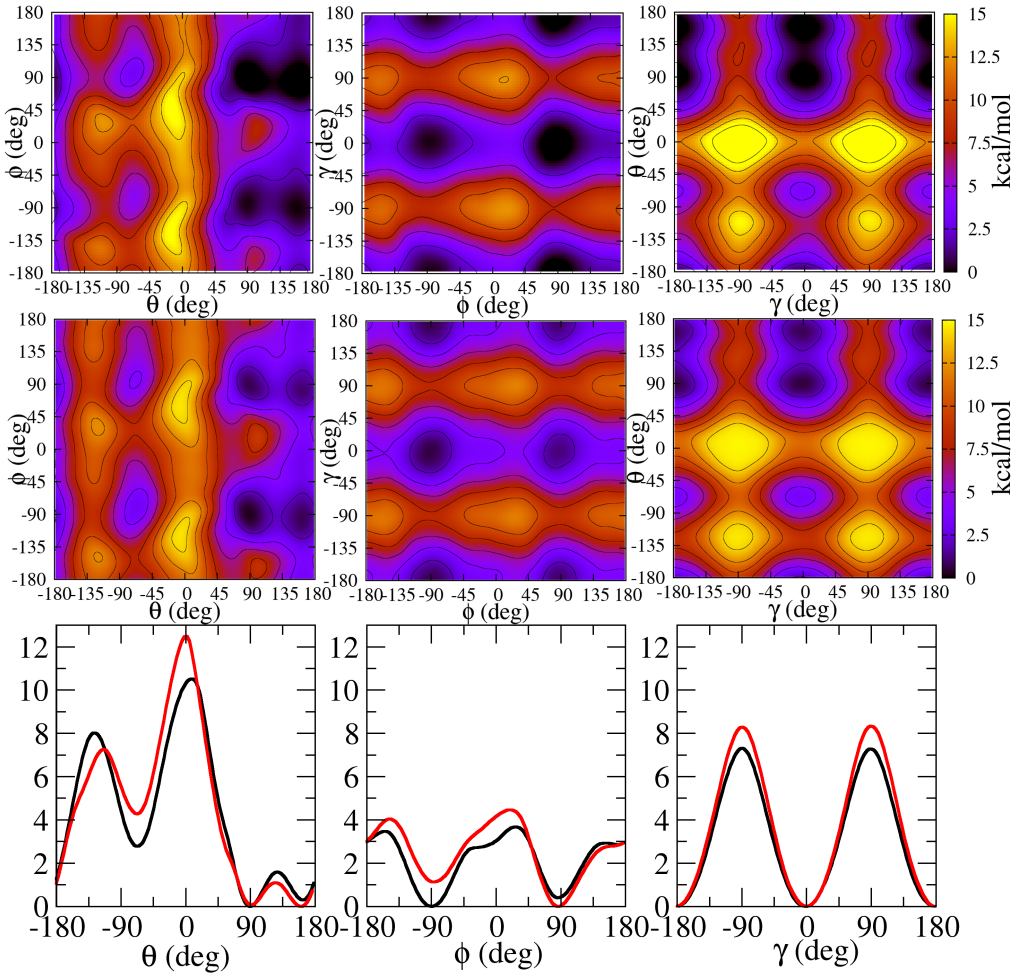

Figure 3.17: Free Energy Surface maps at 300 K as a function of the dihedral collective variables used during metadynamics for ECg. Contour spacing is 2 kcal/mol. The top row is the vacuo metadynamics; the middle row is the solvent metadynamics; the bottom row shows the free energy profile as a function of each collective variable integrated over the other two with the black line indicating the solvent profile and the red line indicating the vacuo profile.

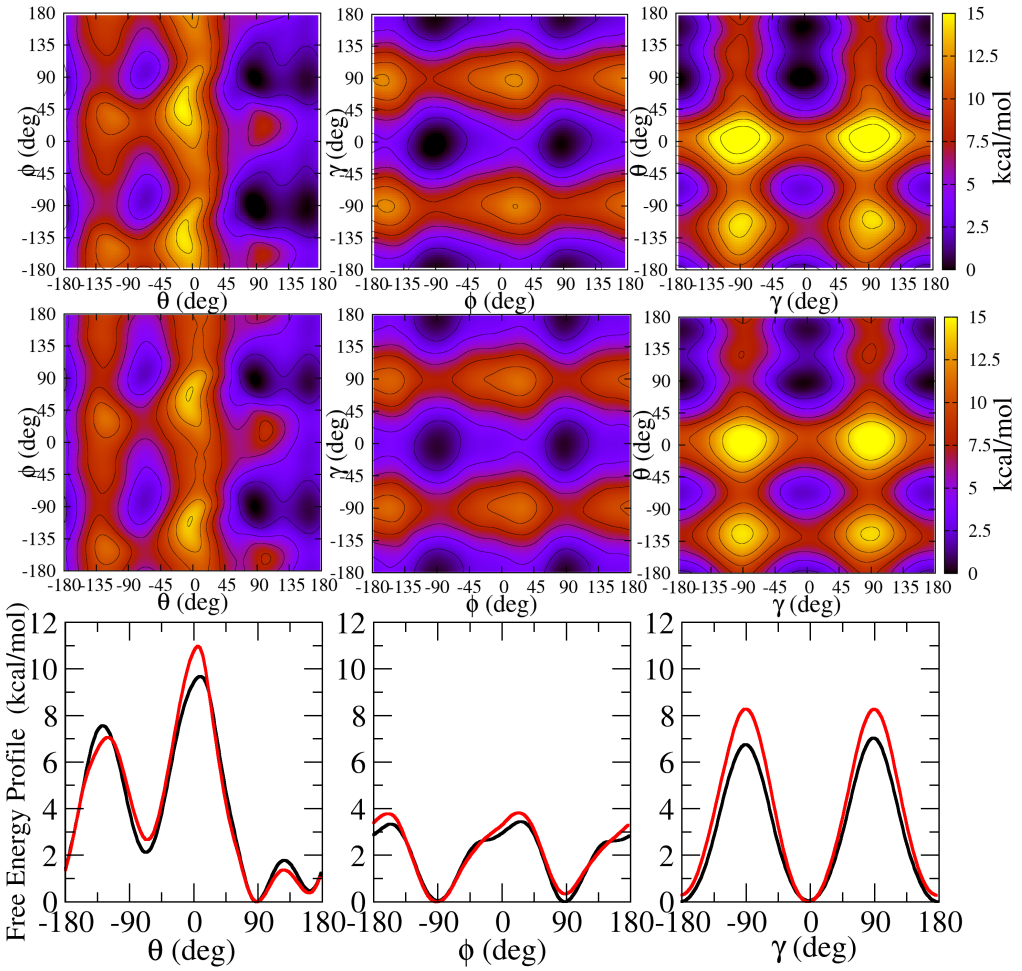

Figure 3.18: Free Energy Surface maps at 300 K as a function of the dihedral collective variables used during metadynamics for EGCmg. Contour spacing is 2 kcal/mol. The top row is the vacuo metadynamics; the middle row is the solvent metadynamics; the bottom row shows the single collective variables projected as a function of the free energy with the black line indicating the solvent profile and the red line indicating the vacuo profile.

depositions. This is demonstrated in Figure 3.20 for the torsional CVs of EGCg in water simulation. As the well-tempering method was used, a converged simulation should have near negligible differences in the profiles of these collective variables after a certain number of profiles.

From inspection of Figure 3.20 it can be seen that the only distinct profiles are those of the first (in black) and the second (in red) with the

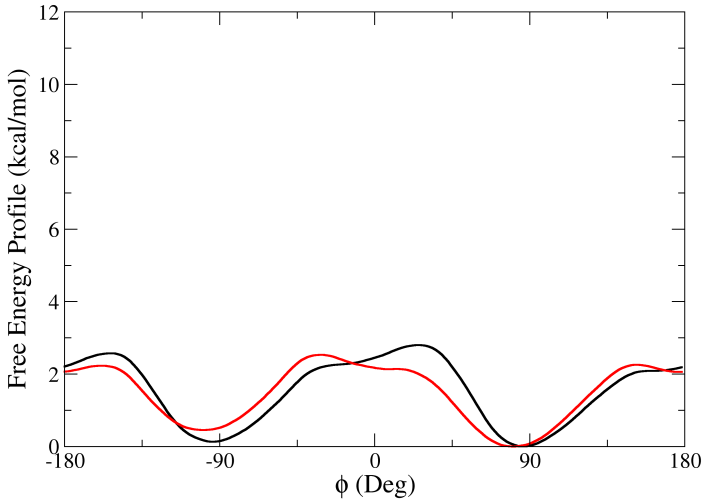

Figure 3.19: Free energy profile of the single EGC metadynamics collective variable, the ϕ torsion as a function of the energy. In black is the solvent profile while in red is the vacuo profile.

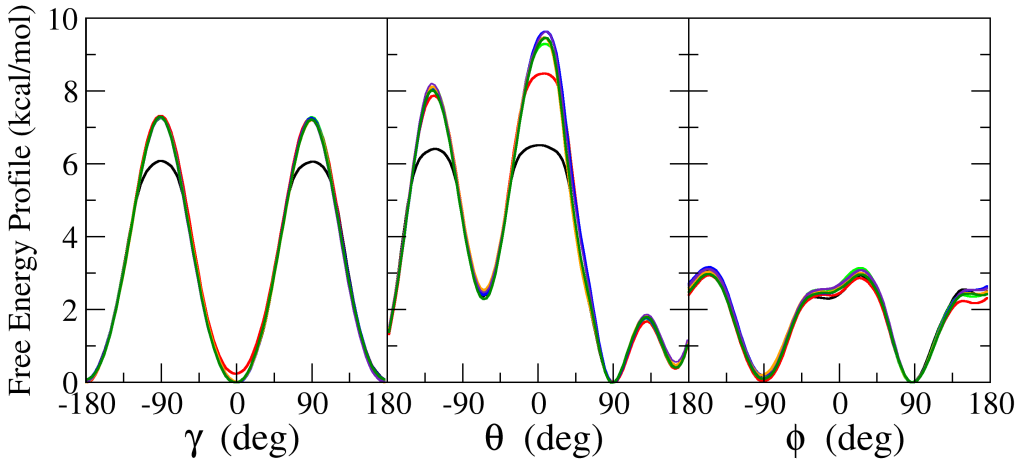

Figure 3.20: Overlaid progression of the torsion collective variables for EGCg in water taken at 10,000 Gaussian increments, setting each profile minima to 0 kcal/mol. Particularly noticeable are the final profile (dark green), the first profile (black) and the second profile (red).

Table 3.2: The standard deviation associated with the solvated catechin metadynamics simulations (kcal/mol).

Profile	Standard Deviation
EGCg γ	0.023
EGCg θ	0.078
EGCg ϕ	0.051
ECg γ	0.058
ECg θ	0.126
ECg ϕ	0.037
EGCmg γ	0.039
EGCmg θ	0.119
EGCmg ϕ	0.027
EGC ϕ	0.079

remainder largely overlapping, indicating a good convergence. The final profile was then compared with the previous 3 profiles in order to give a value for the the fluctuations in the final profiles. This procedure was performed for all the metadynamics simulations and the standard deviation for each profile can be seen in Table 3.2 for the catechins using torsional CVs in water simulations.

In Table 3.2 it can be seen that the error here is very small for all the profiles, so much so as to be considered negligible and so it appears the simulations are well converged. The θ angle appears to have the largest fluctuations by far which is likely because it was the last CV to be fully explored due to its large barrier at 0°. In the case of ECg and EGCmg, conformational restrictions make overcoming this barrier slightly harder and hence these show larger fluctuations than for EGCg. Very similar values were also derived for the simulations in vacuo but are not shown, for brevity.

3.4 Cardiac Drugs: EMD 57033 and Levosimendan

EMD 57033 and Levosimendan are part of a class of drugs which are believed to affect the strength of the cardiac muscle response to intracellular calcium (discussed in Chapter 1) without the need to directly modify the in-

tracellular calcium concentration in the system.[53, 143] This is carried out either by making the regulatory thin-filament proteins more sensitive to the calcium present or by stabilising their calcium induced changes.[144–146] Both EMD 57033 and Levosimendan are displayed in Figure 3.21.

Structurally EMD 57033 has similarities to the catechins in that it has several smaller rings attached to central double ring though the atomic composition is very different. EMD 57033 is known to increase contractile function and has been shown by solution NMR spectroscopy to interact with the C terminal domain of the cardiac protein troponin C in a location similar to that of EGCg.[5, 55]

Levosimendan is known to increase the calcium sensitivity of the cardiac system, thereby increasing cardiac contractility without increasing intracellular calcium. Levosimendan exerts its effect by binding to cardiac troponin C in a calcium-dependent manner. It is also believed to promote smooth muscle relaxation by opening adenosine triphosphate (ATP)-sensitive potassium channels. These combined effects result in an increased force of contraction, as well as reduced muscle stress before and after contraction.[147]

If catechins such as those investigated earlier do indeed exhibit an effect on cardiac contractility as has been suggested, it is worth comparing their interactions with that of a known calcium agent such as EMD 57033 or Levosimendan

Force field parametrisation of both molecules was carried out in a manner described in Section 3.2. The molecules aromatic rings prefer planar orientations and thus RESP charges were derived from the combination of these 180° rotated minima configurations. Simulations in vacuo and water solvent using AMBER (with the GAFF force field) were then undertaken with a protocol similar to that of the catechins and the statistics were collected for 50 ns simulations.

Hydrogen bond analysis was carried out on the trajectories where all nitrogen, oxygen and sulfur atoms were deemed capable of forming H-bonds. For sulfur H-bonds the distance cut-off was increased to 4 Å.[148] The results are seen in Table 3.3. EMD 57033 is entirely unable to self-interact via intra-molecular hydrogen bonding while for Levosimendan it is possible in vacuo between one of the nitrile groups which lies in a planar position

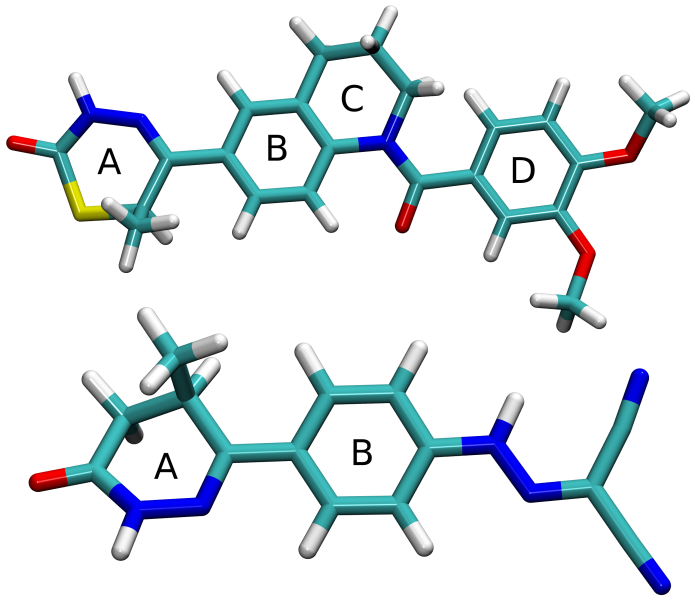

Figure 3.21: Schematic structures of EMD 57033 (top) and Levosimendan (bottom) with labelled rings.

Table 3.3: Total Intra- and Inter-molecular hydrogen bonds during simulations of known calcium sensitisers.

	Intra		Inter
Molecule	Vacuo	Water	Water
EMD 57033	0.0	0.0	6.2
Levosimendan	0.2	0.0	7.9

with the nearby N-H group. The effect of this can be seen in Figure 3.21 as a slight asymmetry between the two nitrile groups. This interaction is not reproduced in water solvent. While Levosimendan is the smaller molecule it appears it can interact more successfully externally via H-bonds on the nitrile groups as shown by the nearly two extra hydrogen bonds, with respect to EMD 57033, formed in water. Though there are a similar number of aromatic rings to the catechins in these molecules, no aromatic stacking can occur due to the relatively planar arrangements of the rings. In the case of EMD 57033 this means that, while it has roughly the same number of atoms as the catechins discussed earlier (EGCg = 51, ECg = 50

EGCmg = 53, EMD 57033 = 53), it is structurally longer which changes its binding and interaction capabilities. Compared to the catechin simulations it is clear that these molecules have a reduced solvent interaction (as we would expect given their molecular composition) and both molecules are likely to favour slightly hydrophobic environments, as is seen with the deep binding of EMD 57033 to the troponin C hydrophobic cleft.[55]

Chapter Conclusions

In this Chapter we have parametrised classical force fields based on GAFF for the green tea catechins EGCg, ECg, EGCmg and EGC as well as the cardiac drugs EMD 57033 and Levosimendan. The quality of these force fields have been assessed with a comparison to first principles calculations. These have then been used for molecular dynamics and metadynamics simulations which have allowed us to evaluate solvation effects and to explore the conformational landscape of the green tea catechins. The force fields will be used in Chapters 5 and 6 to study the interaction of EGCg and the calcium sensitier EMD 57033 with the cardiac muscle troponin C.

Chapter 4

Electronic Properties of Green Tea Catechins

4.1 Chapter Overview

In this Chapter the electronic structure of several radicals of the EGCg and EGC catechins have been interrogated using Hartree Fock, Density Functional Theory and Møller Plesset methods. The importance and role of these radicals and antioxidation properties of the catechins is discussed. Three major mechanisms have been investigated, hydrogen abstraction transfer, single electron transfer and sequential proton loss electron transfer, and the energies for each have been calculated and compared. An analysis of the suitability of widely available functionals in DFT has also been carried out, paying special attention to their ability to produce accurate estimates for the spin density localisation of these types of systems. The use of an implicit polar solvent has been shown to drastically change the relevant energies of cationic and anionic states depending on the external dielectric used.

4.2 The Importance of Catechin Radicals

Catechin radicals are believed to be important when considering their cancer prevention properties due to their interactions with reactive oxidative species; however due to the reactive nature of such molecules it can be difficult to make accurate analysis in experimental conditions. It is therefore

useful to investigate their properties using computational techniques to elucidate the electronic characteristics of the neutral and radical states of green tea catechins.

This Chapter focusses on the antioxidant action of the green tea catechins. The following quantities are evaluated:

Bond-dissociation enthalpy (BDE) - This is defined as the change in enthalpy of a system (at standard conditions) when a covalent bond is broken. In chemistry, BDE is considered in terms of the breaking of one mole of a reactant, in the gas state, into radical products and the average bond breaking energy is usually taken (for example, for each C-H bond in a molecule). In this work the homolytic case is considered, where each product retains one of the electrons involved in the bond. As this work focusses on static calculations of single molecules where a specific bond is broken, the use of BDE here is better thought of as the removal of a hydrogen from the parent molecule to form a neutral radical.

Ionisation Potential (IP) - This is the energy required to remove an electron to an infinite distance from a molecule or atom. In this work, the first ionisation potential is investigated which is the removal of the electron from the highest occupied molecular orbital and is the difference between the energy of the neutral molecule and the cation. Additionally, the adiabatic ionisation potential is investigated here which means that the geometry of the cation has been re-optimised into its new energy minimum position. This is opposed to the vertical ionisation energy where the energy of the cation is calculated for the same structure as the neutral molecule.

Proton transfer - This is where a proton is removed from the molecule, leaving a negatively charged ion. It can be thought of as a calculation of the acidity of a molecule. The definition of a Brønsted-Lowry acid is a substance that can transfer a proton to another substance.[149] This is important in the case of sequential proton-loss electron transfer as the initiation of such reactions is the removal of a proton from the molecule, followed near-instantaneously by the removal of an electron. The important difference between this and BDE is that the proton and electron involved are not from the same bond in the reactant, nor do they attach to the same location in the product and so it is useful to think of these as separate process here.

Lipid peroxidation, a chain propagating reaction in which hydrogen atoms are abstracted and many damaging radicals can be formed, has been shown to lead to serious health problems, especially cancer.[150, 151] The production of these radicals can be prevented through the interception of peroxyl radicals by polyphenol antioxidants (in our case the green tea catechins). Several mechanisms have been proposed for this effect: hydrogen abstraction transfer (HAT); single electron transfer (SET); and sequential proton loss electron transfer (SPLET).

Lipid peroxidation occurs in three major stages, initiation, formation of peroxyl radical and chain propagation:

$$ROO\cdot + LH \rightarrow L\cdot + ROOH \tag{4.1}$$

$$L\cdot + O_2 \rightarrow LOO\cdot \tag{4.2}$$

$$LOO\cdot + LH \rightarrow LOOH + L\cdot \tag{4.3}$$

where we have some unsaturated fatty acids and an initial radical from which alkyl radicals $L\cdot$ form, which in turn go on to form lipid hydroperoxyl radicals $LOO\cdot$. Peroxyl radical formation in Equation 4.2 is very fast ($\sim 10^9 M^{-1}s^{-1}$); however the hydrogen transfer reaction in Equation 4.3 is quite slow ($\sim 10^1 M^{-1}s^{-1}$) which allows polyphenol antioxidants (PhOH) to intercept peroxyl radicals and prevent chain propagation. In the HAT mechanism:

$$LOO\cdot + PhOH \rightarrow LOOH + PhO\cdot \tag{4.4}$$

This reaction is dependent on the bond dissociation enthalpy of the polyphenol antioxidants: the weaker the O-H bond of the hydroxyl group, the faster the reaction in Equation 4.4. The important quantity here is therefore the BDE.

SET reactions can also occur through the following process:

$$PhOH + LOO\cdot \rightarrow PhOH^+ + LOO^- \tag{4.5}$$

$$PhOH^+ + H_2O \leftrightarrow PhO\cdot + H_3O^+ \tag{4.6}$$

$$LOO^- + H_3O^+ \leftrightarrow LOOH + H_2O \tag{4.7}$$

This reaction is solvent dependent, as the important quantity to calculate is the IP of the hydroxyl group, but results in the same products as the previous HAT mechanism.

The final proposed mechanism is SPLET, also known as Proton Coupled Electron Transfer:

$$PhOH \rightarrow PhO^- + H^+ \tag{4.8}$$

$$PhO^- + ROO\cdot \rightarrow PhO\cdot + ROO^- \tag{4.9}$$

$$ROO^- + H^+ \leftrightarrow ROOH \tag{4.10}$$

This can also be thought of as the acidity of the molecule in that it is dependent on the ability of the PhOH to act as a proton donor. This mechanism can lead to false identification of some reactions as HAT rather than SET as SPLET can occur rapidly in some environments.[152]

In many cases SET reactions are slower than those of HAT and this is also believed to be the case for green tea catechins indicating that HAT is the dominant mechanism.[153] Experimental investigations of the products of both mechanisms have shown that the reactions occur on the phenolic B ring of the catechin (ECg, EGCg and EGC were investigated)[153, 154] even if the galloyl moiety (B′ ring) is present. The radical scavenging activity of these molecules has been experimentally predicted to be $ECg \leq EGC < EGCg$ based on their rate of reaction with peroxyl radicals, suggesting that EGCg is of most interest in this area.[155] Computational work on these molecules using a locally dense basis set (LDBS) method has suggested $ECg < EGC \leq EGCg$ on the basis of calculated δBDE values (relative to phenol).[153] The three processes (SET, HAT and SPLET) have the same products so by calculating the energy cost of the limiting step (proton

transfer, electron transfer or hydrogen transfer) it may be possible to gain insights into the energetic favourability of these mechanisms.

4.3 Comparison of Radical Scavenging Mechanisms

The radicals were formed by taking the minimised neutral structures of the catechins tested and removing an electron (to form the cation), a proton (to produce an anion) or a hydrogen (to form a charge-neutral radical). Two of the catechins investigated in Chapter 3 were chosen to investigate the quantities of interest. These were EGCg, due to its believed superior radical scavenging ability and larger abundancy, and EGC due to its wide structural divergence compared to the others due to the lack of the B ring. In addition, both these molecules were considered to be the most efficacious in radical scavenging ability in previous work.[151, 153, 154, 156] The selected catechins were optimised in neutral, cationic, anionic and $PhO\cdot$ forms as it was assumed that the radicals would have enough time to re-orient themselves into favourable conformations before further chemical processes occur. Two neutral radicals produced via HAT mechanisms were used, the radical with an external hydroxyl hydrogen removed (H8) and the radical with the internal hydroxyl hydrogen removed (H7), both from the B ring. These are shown in Figure 4.3, while the structures used for EGC were very similar (see Figure 3.1) and are labelled in the same manner.

They were chosen because experimental work has implicated the hydroxyl groups on this ring as being overwhelmingly targeted for these interactions over other locations.[156] In the anionic case, the initial structures are exactly the same as for the neutral radicals but with only a proton removed at the H7 or H8 positions leaving a -1 charged singlet state. The H8 radical was considered indistinguishable from the H6 form as the only difference lies in the dihedral ϕ rotation of $\pm 90°$ which has extremely small barriers to rotation, as found in Chapter 3. In all cases a basis set of 6-311++G(2d,2p) was used in order to attempt to precisely capture the polarisation effects of the cationic and anionic states. A selection of functionals

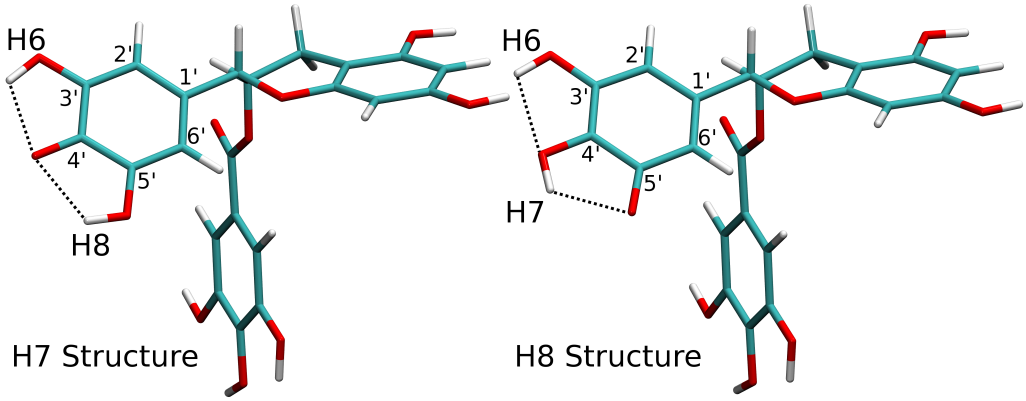

Figure 4.1: Initial structures of the EGCg neutral and anionic radicals. Left: the B ring interior hydroxyl hydrogen is removed from the 4' carbon position (labelled H7). Right: the exterior hydroxyl hydrogen is removed from the 5' carbon position (labelled H8). Both initial structures were modified by rotating their hydroxyl groups into optimal intra-molecular hydrogen bonding positions before geometric optimisation. Carbon B ring labels are also shown.

were used to compare the applicability of several widely used methods with radical systems. The calculations were carried out with HF as well as a selection of DFT functionals to compare their performance, including B3LYP, CAM-B3LYP and M05-2x: B3LYP[157] was chosen as it is a popular hybrid functional and widely tested which has shown itself to be useful for biological molecules; CAM-B3LYP[82] was included to test whether improvement could be found with the Coulomb attenuating method to correct long-range interactions; M05-2x[158, 159] was used as it was parametrised for non-metals with double the amount of long range exchange (hence 2x) and was shown to perform very well for radical reaction barrier heights as well as non-covalent interactions. These were also repeated using the Polarisable Continuum Model (PCM) to reproduce an aqueous environment which is important to account for polarisation effects using a dielectric constant of $\varepsilon = 78.36$ for water.[160] Zero point energy corrections were included from frequency calculations of the geometrically optimised stuctures. For each functional tested the radical was compared with the respective neutral molecule to calculate the relevent adiabatic proton transfer, IP or BDE. In the case of the BDE the energy difference was modified by the functional

Table 4.1: IP, BDE and proton dissociation energies in kcal/mol for EGCg in gas phase (top) and in PCM implicit water solvent (bottom) including zero point energy and thermal corrections to room temperature.

Functional	BDE		H^+ transfer		IP
	H7	H8	H7	H8	
HF	44.21	46.36	336.12	339.15	149.80
B3LYP	67.86	74.79	315.78	321.21	164.13
CAM-B3LYP	68.85	76.79	316.21	321.28	175.39
M05-2x	73.54	80.95	329.95	320.23	177.59
PCM-H_2O	-	-	-	-	-
HF	45.44	47.67	290.51	291.68	112.37
B3LYP	69.39	75.32	272.66	275.88	136.47
CAM-B3LYP	71.16	77.44	272.01	275.23	137.84
M05-2x	74.75	80.74	286.36	287.20	141.42

Table 4.2: IP, BDE and proton dissociation energies in kcal/mol for EGC in gas phase (top) and in PCM implicit water solvent (bottom) including zero point energy and thermal corrections to room temperature.

Functional	BDE		H^+ transfer		IP
	H7	H8	H7	H8	
HF	43.69	46.30	334.92	336.56	148.52
B3LYP	67.70	74.57	316.88	320.75	168.10
CAM-B3LYP	69.58	76.47	316.98	320.52	173.86
M05-2x	72.39	79.66	316.75	320.37	176.42
PCM-H_2O	-	-	-	-	-
HF	45.39	47.71	292.68	292.89	111.44
B3LYP	68.77	74.75	277.22	280.04	135.03
CAM-B3LYP	71.17	77.10	276.80	279.13	137.11
M05-2x	74.13	80.67	276.49	278.15	140.88

energy of the hydrogen atom of -0.4998 Hartrees for HF, -0.5023 Hartrees for B3LYP and -0.4989 Hartrees for CAM-B3LYP and -0.4993 Hartrees for M05-2x. The results are shown in Table 4.1 for EGCg and Table 4.2 for EGC.

In both Table 4.1 and 4.2 one can see a clear ranking of energy for the quantities measured. It should be noted that these calculations cannot give a complete picture as radical chemistry is highly dynamic and there are

many other factors involved that cannot be easily modelled with static calculations. Energy barriers are also not evaluated. However, for all functionals tested it appears that the BDE in the HAT mechanism is significantly smaller than the IP in SET and the H^+ transfer in SPLET. It is also clear that the central, H7, abstraction would be more favourable than the H8, though the differences in energy are not large and it may be possible for either to occur. Compared to data reported by Leopoldini et. al.[151] we can see that EGCg and EGC compare favourably against Epicatechin and Catechin (catechins similar to EGC and EGCg also found in green tea) in both IP and BDE values in the gas phase (reported as Epicatechin BDE - 73.7 kcal/mol, IP - 170.8 kcal/mol; Catechin BDE - 74.2 kcal/mol, IP - 169.7 kcal/mol. All calculations using B3LYP with a 6-311++G(d,p) basis set in Gaussian). We immediately notice a stark reduction in IP and H^+ transfer energy when including the effects of water solvent which is not reproduced in the BDE values. In fact these BDEs seem slightly larger under PCM conditions, though still smaller than the other quantities.

The H^+ transfer energy in both catechins appears quite high, in large part due to the inability of these methods to calculate the energy of isolated protons, even in water. It appears unlikely that proton transfer would be favoured over ionisation or hydrogen abstraction, given these results. Comparison of the results shown in the two tables suggest that for these mechanisms $EGC \approx EGCg$ consistent with previous computational work.[153]. While EGC appears to have slightly lower computed BDE and IP energies in general than EGCg, the differences are smaller than 1 kcal/mol especially with the BDE values there is almost no difference in energy. Earlier computational work using the LDBS method with B3LYP provided a reference value for ΔBDE and ΔIP of -16.1 kcal/mol and -27.9 kcal/mol respectively where the values are given relative to those of phenol and are in the gas phase. The Table 4.1 value for B3LYP in gas phase corresponds to -14.0 and -26.9 kcal/mol in good agreement, though this changes significantly for the other functionals tested and becomes ΔBDE of -11.5 and ΔIP of -6.8 when implicit solvation is included. Various experimental works have placed the BDE value of phenol at 88.7 kcal/mol in the gas phase and 85.2 kcal/mol in water solution.[161] Previous DFT computational works men-

tioned however place these value at 87.1 kcal/mol in the gas phase using a Restricted open-shell approach[153] and 82.9 kcal/mol in gas phase and 97.1 kcal/mol in water, using PCM.[156] Both calculations used the B3LYP functional and so to compare with our own we find value of 82.1 kcal/mol in gas phase and 81.1 kcal/mol in water solution. While our value is further from experimental estimates than the restricted open-shell approach in gas phase, we do see the same experimental trend of a reduction in the BDE from gas phase to solution unlike the work of Leopoldini et al.[156] who give a solution BDE over 10 kcal/mol too high. A possible reason for this discrepancy is that in this earlier work the structures were optimised using AM1 methods with a 6-31G basis set.

4.4 Spin Density Distributions

In order to estimate the accuracy of the calculations performed it is useful to examine the localisation of electron spin in the systems. As previously discussed, experimentally it is believed that radical processes are carried out on the pyrogallol (B) ring of these catechins. We would therefore expect the excess unpaired spin in the cationic and neutral systems to be localised here. This is in fact an extremely important test of the applicability of the DFT functional as the methodology is completely dependent on the electron (and therefore spin) density. The location of the spin density of these radical states is also likely to be where electrophilic attack occurs on these molecules. Experimental work in deriving the products of such radical interactions have overwhelmingly claimed that the site of electrophilic attack is on the catechin B ring.[150, 154, 162, 163] Previous computational work[151, 156] assert that the unpaired spin of the radicals of Catechin and Epicatechin are delocalised over the B and C rings in gas phase which appears slightly at odds with the experimental view that the B ring alone is the crucial site. The spin densities of selected neutral and cationic EGCg radicals are shown in Figures 4.2, 4.3 and 4.4. EGC radicals are shown in Figures 4.5, 4.6 and 4.7.

For all neutral radicals tested, of both EGC and EGCg, the observed B3LYP spin density is highly similar to the CAM-B3LYP and M05-2x func-

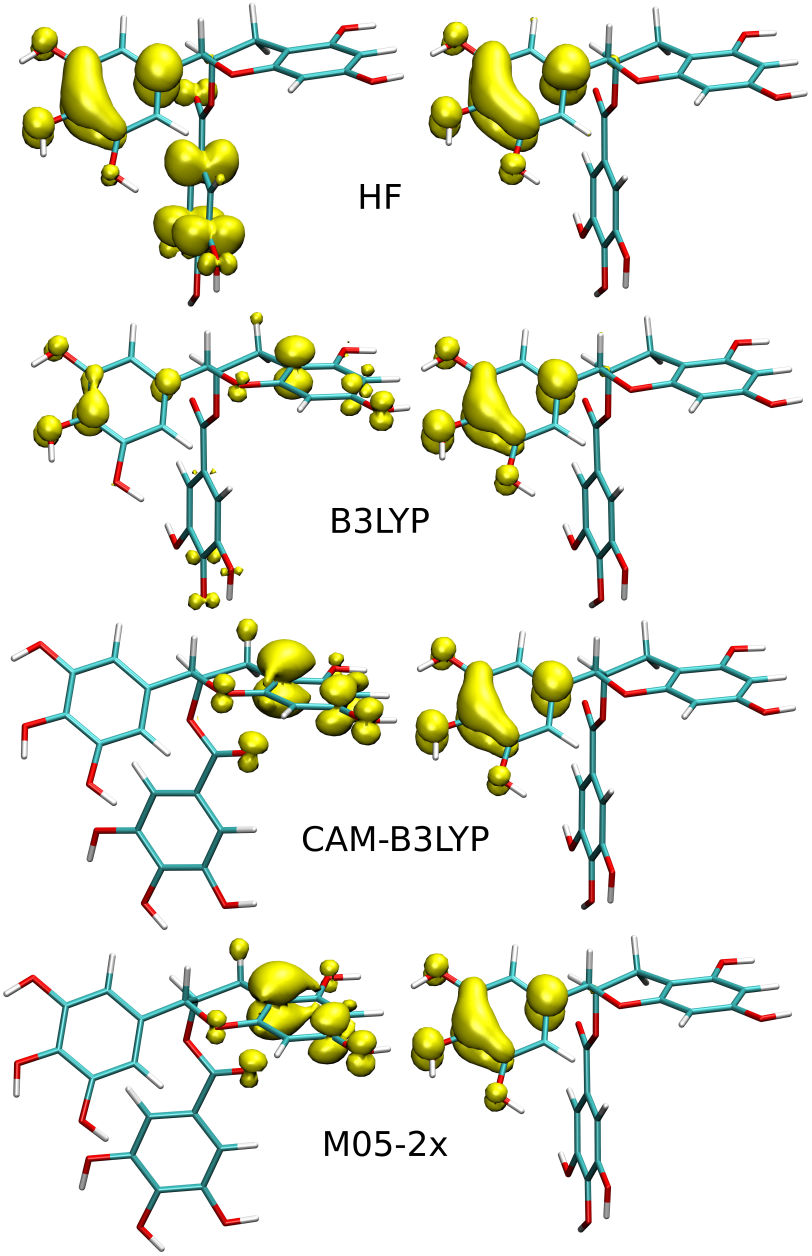

Figure 4.2: Cation spin densities for the EGCg molecule at fixed isovalue 0.005 electrons per Å^3. Left: gas phase optimisations. Right: PCM water solvent optimisations.

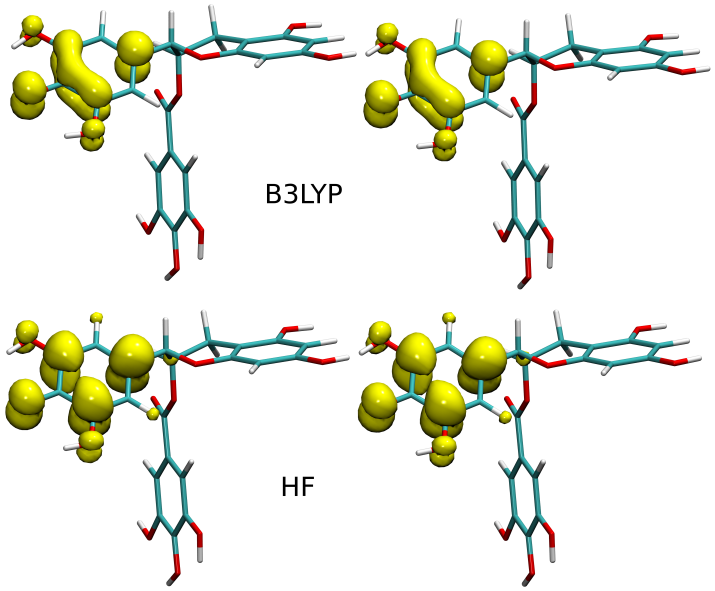

Figure 4.3: Selected neutral H7 EGCg radical spin densities at fixed isovalue 0.005 electrons per Å^3. Left: gas phase optimisations. Right: PCM water solvent optimisations.

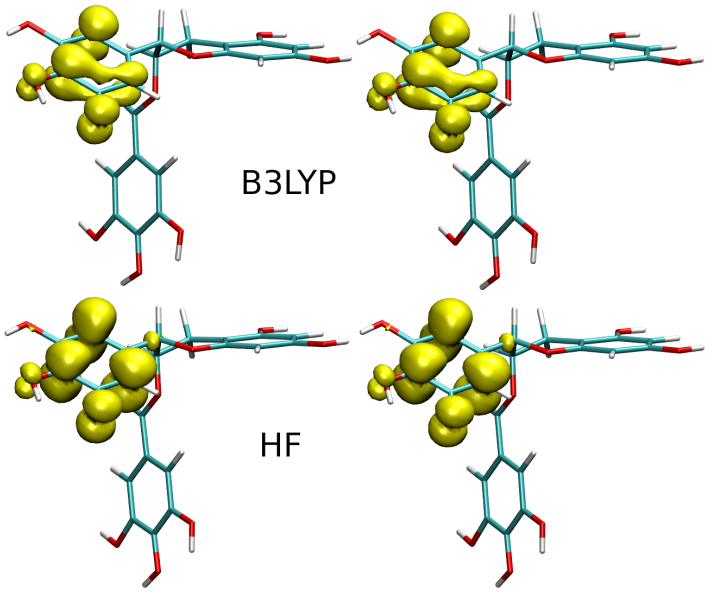

Figure 4.4: Selected neutral H8 EGCg radical spin densities at fixed isovalue 0.005 electrons per Å^3. Left: gas phase optimisations. Right: PCM water solvent optimisations.

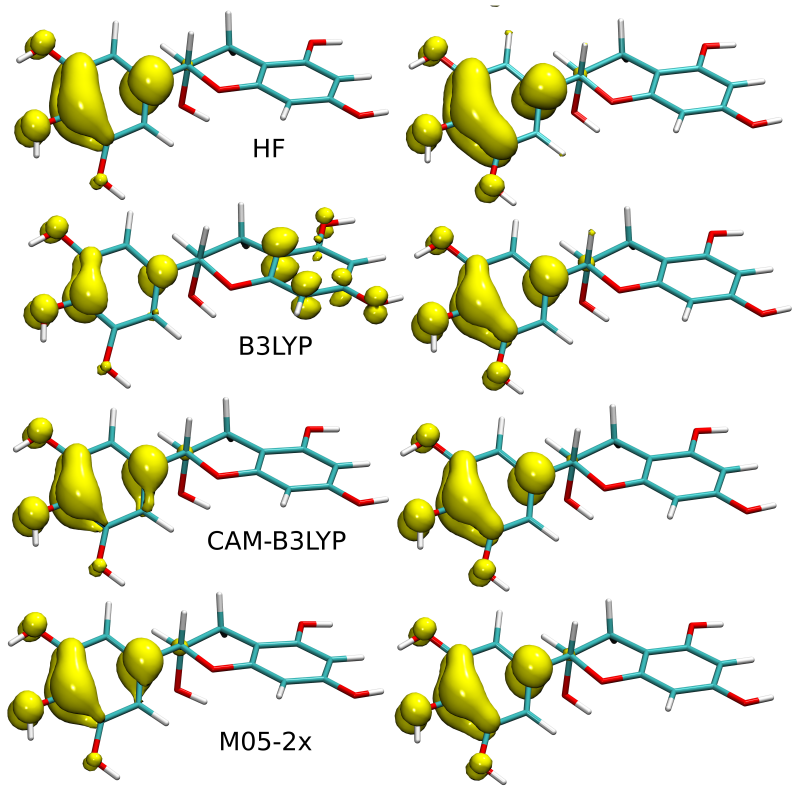

Figure 4.5: Cation spin densities for the EGC molecule at fixed isovalue 0.005 electrons per Å^3. Left: gas phase optimisations. Right: PCM water solvent optimisations.

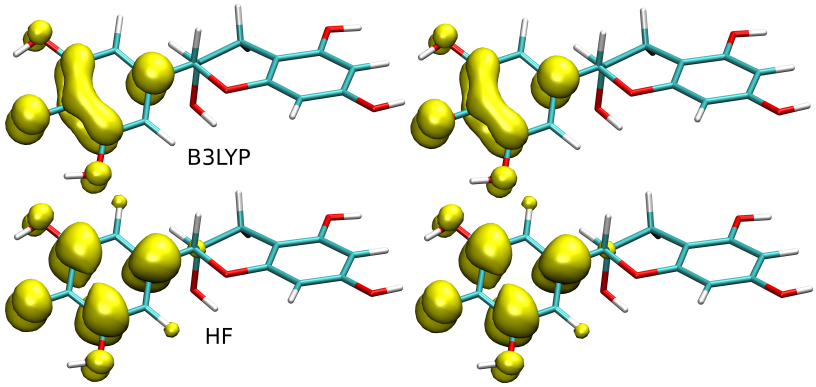

Figure 4.6: Selected neutral H7 radical spin densities are shown at fixed isovalue 0.005 electrons per Å^3. Left: gas phase optimisations. Right: PCM water solvent optimisations.

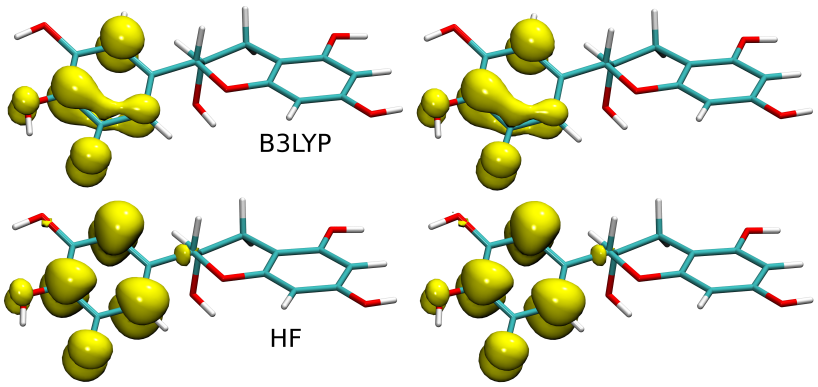

Figure 4.7: Selected neutral H8 radical spin densities are shown at fixed isovalue 0.005 electrons per Å^3. Left: gas phase optimisations. Right: PCM water solvent optimisations.

tionals as no qualitative differences between them can be seen in either gas phase or aqueous solution. The cation calculations of both catechins in gas phase give highly varied spin density distributions, depending upon the chosen functional. In addition, as can be seen in Figure 4.2, the optimised structural conformation of EGCg is also quite different between the functional chosen, primarily in the rotation of the θ dihedral. Though this does not appear to affect the trend of IPs in Table 4.1, the B3LYP functional indicates a spin density spread over the entire molecule in the cationic state and it is unlikely that this is an accurate picture of the spin density as the radical reactions are believed to centre on the B ring in all cases. In the case of the CAM-B3LYP and M05-2x the structure is unable to optimise near the neutral conformation and there is considerable rotation of the dihedral θ angle with spin localisation on the A ring. This rotation enables interaction between the A ring and carbonyl oxygen from the B′, hence there is also a small accumulation of spin denisty on this oxygen. For the HF calculation the spin density appears spread between the B and B′ rings in a symmetric manner around the rings. In addition it was found that spin contamination is significant in the HF calculations, the annihilated S^2 value being 0.87 for the cation and between 1.01 and 1.02 for the neutral radicals. For every functional tested the effect of PCM solvent with a high dielectric constant drastically changes the distribution of spin density by

restricting it entirely onto the B ring of the molecule and localised around the 4′ carbon of the ring which corroborates the earlier work of experimentalists discussed previously. With PCM solvent the optimised geometry of the EGCg structure also reverts more closely to the initial neutral orientation and that seen with B3LYP in the gas phase, suggesting that this functional is a reasonable choice for geometry optimisation even if it cannot capture electron spin behaviour. For both neutral radical states the DFT functionals chosen gave almost identical distributions of spin density concentrated, as expected, around the B ring. In the H7 case this leads to spin localisation around the central 4′ carbon bonded to the de-hydrogenated oxygen with highly symmetric spin density around the neighbouring carbon bonds. In the H8 case there are accumulations of spin on the carbons either side of the 5′ carbon with only a small distribution extending over this carbon itself. This distribution is highly asymmetric and appears to indicate that the 4′ carbon is generally preferred and therefore the likely site of electrophilic attack. The Hartree Fock calculations also appear to give reasonably close distributions to the DFT results, but show spin densities symmetrically localised completely over the de-hydrogenated oxygen and alternating carbons on the ring.

To test the effect of the PCM solvent with a small dielectric the B3LYP optimisation of both EGCg and EGC were re-run in benzene ($\varepsilon = 2.27$) and the spin densities of the cation and hydrogen abstracted radicals are shown in Figure 4.8.

The associated IP, BDE and proton transfer energy are shown in Table 4.3 where we can seen similar trends between catechins as was seen in the gas and water calculations.

The hydrogen abstraction radicals show little difference in BDE from that of the earlier gas phase and water calculations but we do see a change in the cation and anion results as we would expect from the results in Tables 4.1 and 4.2. The distributions of spin density is also similar with the neutral radicals showing no real difference to the gas phase results while the cation is considerably closer to the gas phase distribution than the water, though in the case of EGCg no spin density is seen on the B′ ring here. These results seem to indicate that as the external dielectric is increased, the

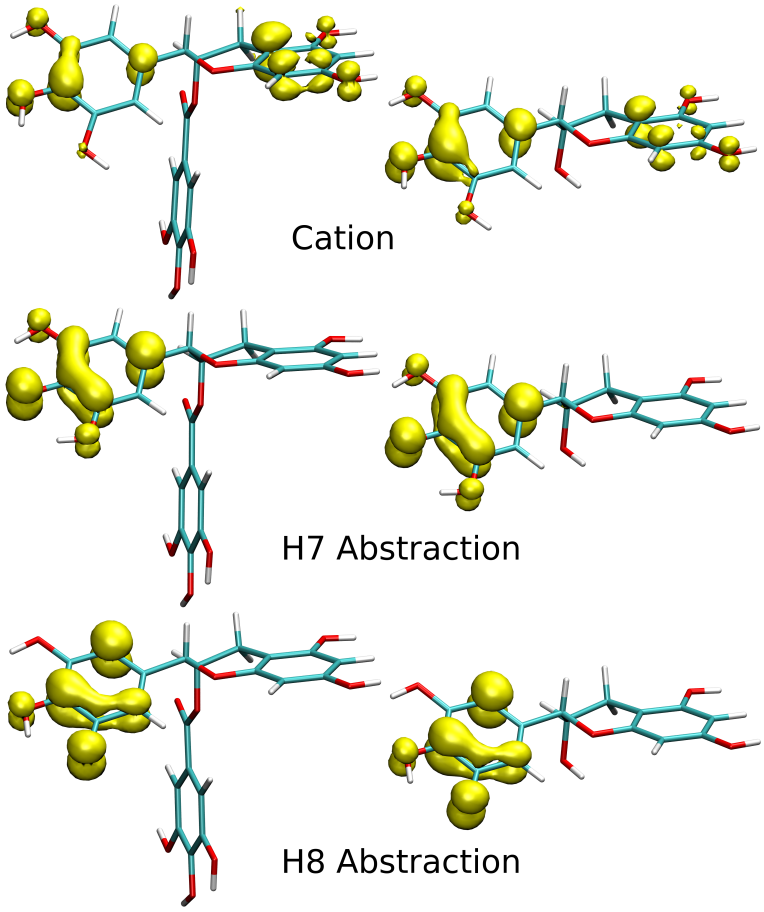

Figure 4.8: Radical optimisations of the EGCg (Left) and EGC (Right) molecules displaying unpaired spin density with B3LYP using benzene implicit solvent at fixed isovalue 0.005 electrons per Å^3.

Table 4.3: IP, BDE and proton dissociation energies in kcal/mol of the radical forms of EGCg and EGC in PCM implicit benzene solvent including zero point energy and thermal corrections to room temperature.

Radical	EGCg	EGC
Cation	148.22	150.23
H7 Abstracted	68.81	67.87
H8 Abstracted	75.33	74.54
H7 Anion	303.12	283.80
H8 Anion	310.38	306.32

spin density of the charged radical localises more around the B ring, giving results closer to that expected from experiment. Overall, DFT calculations of these cationic radicals do not appear to be able to give a reliable spin density without the inclusion of implicit solvent with a high dielectric.

4.5 Møller Plesset Calculations of Cations

Further analysis of the two catechins was undertaken using unrestricted MP2 (UMP2) single point energy calculations on the B3LYP optimised structure. This structure was chosen as it has been shown to give a very similar estimate for the geometry of the minimised system to the PCM results despite poorly resolving the spin density in the gas phase. Energy evaluations were carried out to establish whether the effects of spin contamination play a significant role in the energy evaluations of the system. The structures of the UMP2 with spin density iso-surface can be seen in Figure 4.9.

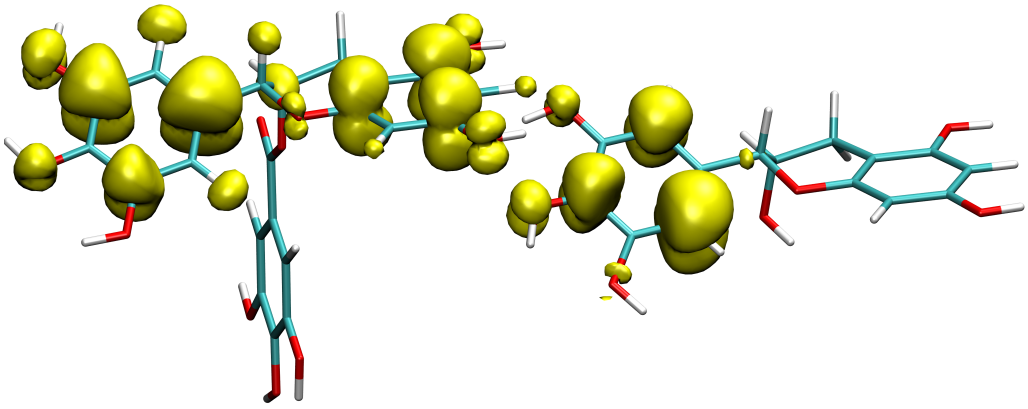

Figure 4.9: EGCg and EGC optimised cation with UMP2 single point energy evaluation displaying the MP2 spin density at fixed isovalue 0.005 electrons per Å^3. In this case the EGCg MP2 energy was slightly lower in the ϕ -90° position while for EGC this remained in the +90°.

Similar to the HF evaluation seen earlier, for EGCg the UMP2 result places concentrations of spin density over alternating carbons 1′, 3′ and 5′ of the B ring but unlike the HF this is also seen on the benzopyran A/C ring instead of the B′ ring. The EGC spin did localise on the B ring but, as with

EGCg, the spin appears to localise over individual B ring carbons, in this case the 2′, 4′ and 6′. Spin contamination was a factor in the unrestricted EGCg case with S^2 found to be 0.781 after annihilation. As this was within 10% of the true value of 0.75 we consider this not to be a dominant effect and is much reduced over the HF values. For EGC the S^2 was 0.757 and so considered negligible. The IP value of the UMP2 EGCg cation was 184.36 kcal/mol while for the EGC this was 187.01 kcal/mol. Both values lie closer to that calculated for the CAM-B3LYP and M05-2x functionals in gas phase earlier and it is interesting to note that these functionals too showed spin density accumulation on the benzopyran ring in the EGCg case though this does not satifactorily explain the EGC cation energy. Overall neither structure appears to represent a better picture of the catechin cation than those calculated with DFT methods earlier despite the increased cost of calculation.

4.6 Chapter Conclusions

Investigation into the IP, BDE and proton transfer energies of the antioxidants EGCg and EGC led to several trends being observed: directly comparing between catechins suggests that $EGC \approx EGCg$, while the quantities themselves rank in terms of energy cost $BDE < IP \ll H^+transfer$ and in both neutral and anionic abstraction cases the central H atom is preferred. Despite reasonable agreement with previous work, analysis of the spin density of the cationic states appears to suggest that the functionals selected led to overly de-localised spin configuration without the use of PCM implicit solvent with a high dielectric constant. In addition there are significant torsional structural differences in the EGCg molecule in the gas phase using the CAM-B3LYP and M05-2x functionals which are very different from the aqueous case. HF calculations in cation and neutral radical states led to significant spin contamination which is only partly improved by Møller Plesset methods and it is clear that these calculations cannot accurately capture the radical characteristics of these catechins.

Chapter 5

Interaction of Epigallocatechin 3-Gallate with Cardiac Troponin C

Chapter Overview

In this Chapter we examine the protein C-terminal troponin C and its interaction with the green tea catechin EGCg. Previous experimental research on these systems is discussed and used to establish rigorous computational protocols for docking and classical simulations which we then analyse and compare to experiment. The suitability of computational docking with regards to this and similar protein systems is investigated and the effect of EGCg on calcium mobility in wild-type and mutated protein is presented. Following this, metadynamics to elucidate the binding site of EGCg is carried out and compared with the work in Chapter 3. The results contained in this Chapter have been published in PLOS ONE.[164]

5.1 Troponin C: Function and Mutation

Troponin is a protein which exists in both cardiac and skeletal muscle and plays different functions depending on the location. Of interest in this work is cardiac troponin (cTn), shown in Figure 5.1, which has a critical role in

regulating relaxation and contraction in the heart. The protein is attached to tropomyosin and lies between actin filaments found in muscle tissue. It is composed of three distinct protein sub-units which each play important roles in this regulatory function. Troponin C (cTnC) binds Ca^{2+} and regulates the thin filament activation, troponin I (cTnI) inhibits contraction in the absence of Ca^{2+} and troponin T (cTnT) attaches the troponin complex to tropomyosin on the actin filament.[165]

As demonstrated by solution NMR[166], troponin C (see Figure 5.2) is a dumbbell shaped protein consisting of two domains comprising the C and N terminals connected by an α-helical linker. Each domain contains two EF-hand helix-loop-helix motifs which act to bind divalent cations, though in the N domain one of these sites is inactive. The hydrophobic core of the regulatory N-terminal (cNTnC) domain is buried inside the protein but unlike in the case of skeletal TnC, which exposes its hydrophobic core when bound to Ca^{2+}, the regulatory domain is not thought to undergo major conformational changes upon binding Ca^{2+}.[4, 166–169] The Ca^{2+} binding sites on the C terminal cTnC are of high affinity and without the ions bound, the domain remains unstructured.[170]

Calcium ion binding is known to be the physiological activator of contractile proteins and so troponin C is of particular interest to researchers looking for ways to modulate heart contraction as the cTnC protein is a useful target for drugs which attempt to either sensitise or desensitise the ability of the system to bind Ca^{2+} in order to restore normal heart function. These types of drugs are considered safer than others currently used such as milrinone[171] which alter the cytosolic Ca^{2+} homoeostasis and can cause arrhythmia or death with continued use. As the regulatory domain of cTnC undergoes the initial binding of calcium ions which triggers the relaxation-contraction cycle, it has traditionally been the target for Ca^{2+} sensitisers/desensitisers; however the structural domain (cCTnC) has received scientific interest of late as its location is the primary binding site of cTnC with the rest of the thin filament and so cCTnC my be a suitable target for drugs which alter the calcium sensitivity of the protein. The switch region of cTnI (residues 147-163) interacts with cNTnC to draw the inhibitory and C terminal regions of cTnI (residues 128-147 and 163-

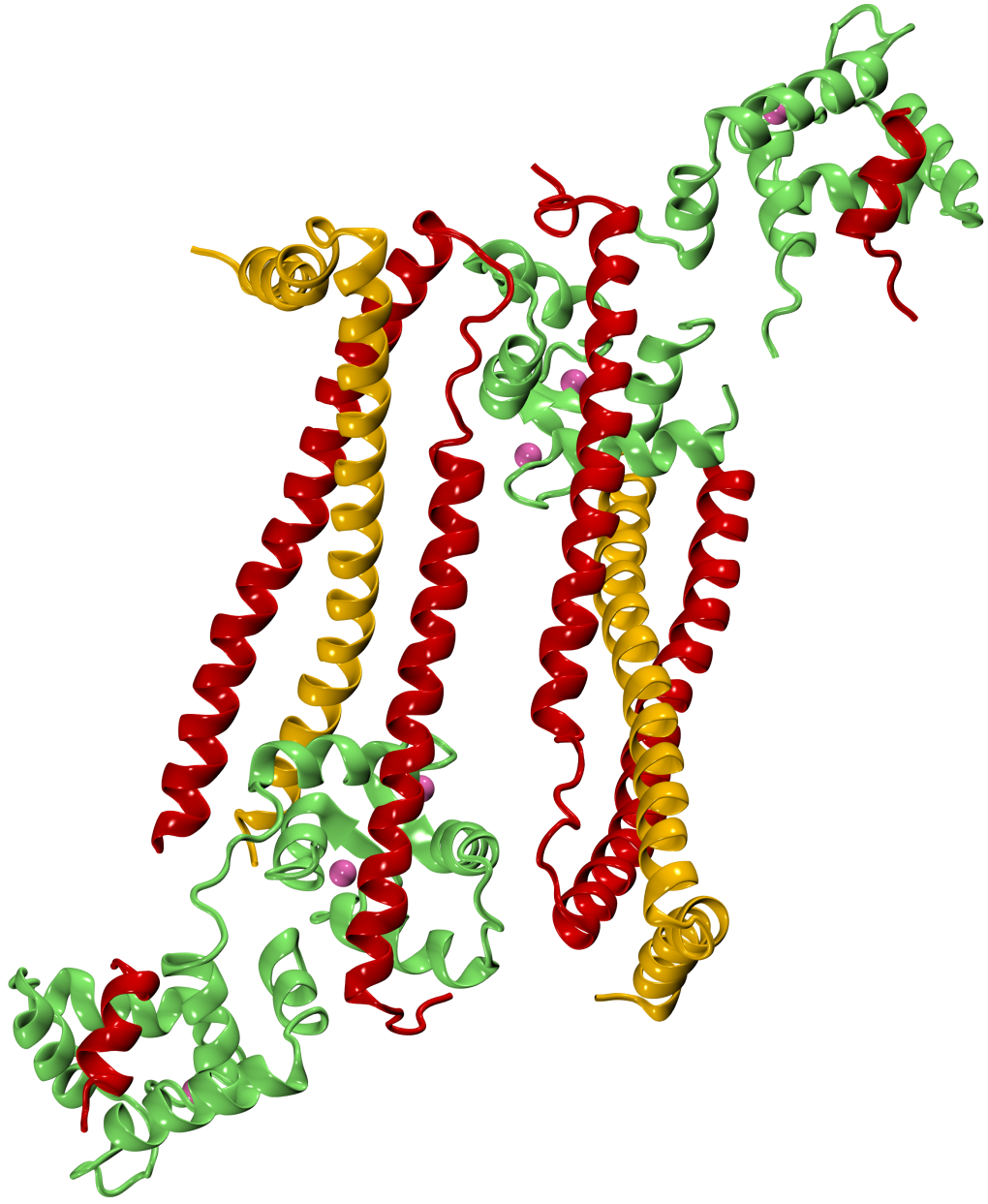

Figure 5.1: X-ray crystal structure of the partially resolved core domain of two entwined human cardiac troponin proteins saturated with Ca^{2+}(in pink). Troponin C (shown in Green), troponin T (shown in yellow) and troponin I (shown in red). PDB code 1J1D[4]

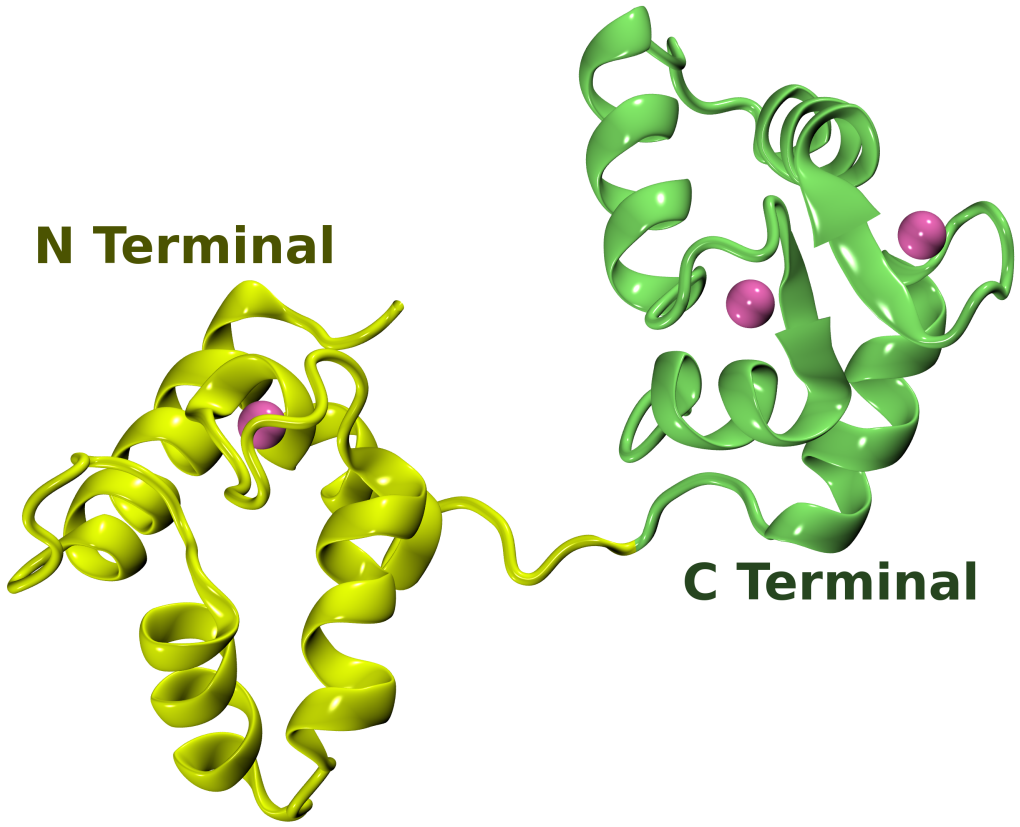

Figure 5.2: View of the dumbbell shaped Troponin C protein sub-unit taken from PDB 1J1D. The N terminal is in light green while the C terminal is in dark green. Three calcium ions are present in the high and low affinity active sites and are in pink.[4]

210 respectively) away from actin during the relaxation-contraction cycle, while the anchoring region of cTnI (residues 34-71) interacts with cCTnC to tether the cTnC protein to the thin filament. The calcium sensitiser EMD 57033 is believed to work by weakening the interaction of cTnI(34-71) with cCTnC and thereby increase the tendancy for cCTnC and cTnI(128-147) to bind.[55] As both EMD 57033 and EGCg are believed to bind near the hydrophobic cleft of cCTnC, it is possible that these have a similar effect. Recent experiments have identified the green tea catechin EGCg as a possible calcium desensitiser which may be of use as a treatment for Hypertrophic Cardiomyopathy[28, 172] however further investigation into these calcium modulators is required to understand precisely the mechanisms which allow

them to carry out such function.[6] This leads to the need for in-depth structural and conformational analyses using atomistic computational methods.

5.2 Rigid and Flexible Docking Studies

The initial research by Sykes et al.[6] provided a useful basis from which it becomes possible to explore and investigate the effects of catechins in a cardiac system. Using the AUTODOCK[127] program suite it was possible to perform a computational docking to compare with the experimental results as well as search for possible alternative binding sites. Several docking calculations were undertaken to establish the viability of the scoring and search algorithm to reproduce the known system (deposited in the pdb database, code 2KDH[6]). The lowest energy NMR structure from the 2KDH pdb is diplayed in Figure 5.3. On the left cCTnC·EGCg with labelled helices and bound calcium ions is shown while on the right cCTnC·EGCg is shown with the protein surface (displayed by residue type) which surrounds EGCg highlighting the highly hydrophobic environment of the cleft.

The preliminary docking of EGCg to cCTnC was carried out with no *a priori* knowledge of the binding site assumed. This gives a quick binding result as bond rotations and translations are not accounted for but tends to provide lower quality bound structures. An all rigid receptor was used in the form of the cCTnC extracted from the lowest energy solution NMR structure with non-polar hydrogens removed. A set of cubic atomic grids which spanned the entire protein as well as a 10 Å buffer region on all sides was included in order to allow translation and rotation of the catechin by the search algorithm. Default Gasteiger charges were applied to the protein with the calculated DFT charges for EGCg (as discussed in Chapter 3. All non-polar hydrogens the ligand were removed and the ligand bonds comprising C-C and C-O atoms which were not part of a ring system were specified to be flexible. The results of this preliminary docking overwhelmingly centred around the hydrophbic cleft in excellent agreement with the established experimental work and so a more refined, flexible docking was carried out. The grid was localised around the hydrophobic cleft with dimensions $94 \times 106 \times 96$ Å in order to improve the clustering of results. A grid spacing

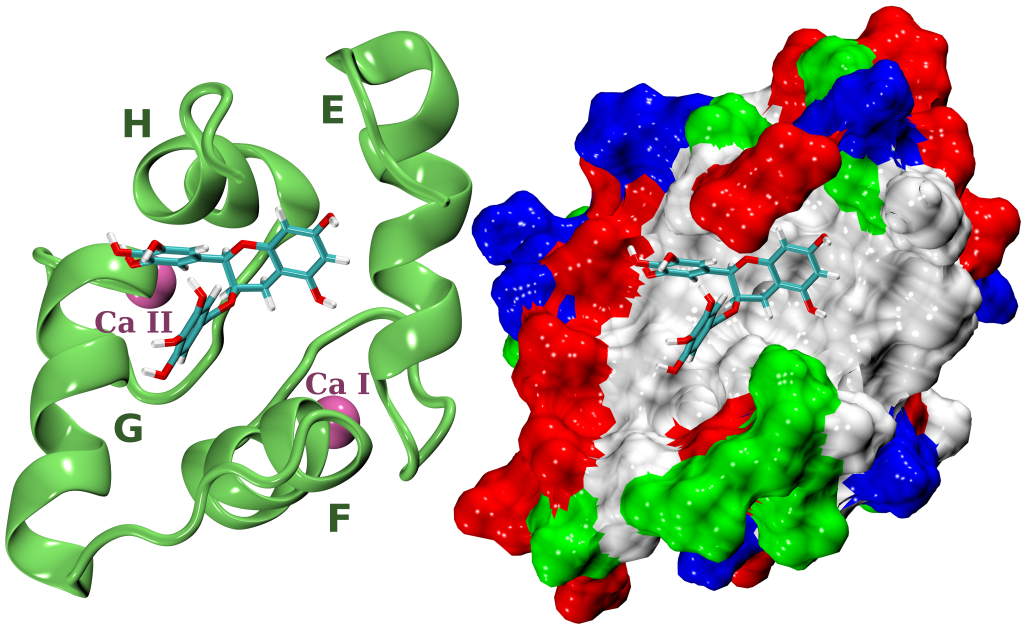

Figure 5.3: 2KDH pdb stucture showing C terminal troponin C in complex with EGCg. Left: Protein-ligand structure shown with labelled helices and calcium ions. Right: Display of protein molecular surface centred on EGCg. Surface is coloured by residue type where white is hydrophobic, red is acidic, blue is basic and green is polar uncharged.

of 0.182 Å^3 was used. Several residues inside or very near the hydrophobic cleft were found to have strong NOE contacts with EGCg experimentally. These were the residues MET120, LEU121, LEU136, MET157 and VAL161 which are displayed in orange in Figure 5.5. Using this knowledge as a basis it is possible to define these residues as 'flexible' to the AUTODOCK program, effectively treating the bonds contained within this selection as able to move in the same manner as the ligand. This helps provide a more refined docking result as the ligand is able to fit more closely to these areas of the protein (if it is favourable to do so). 200 Lamarckian genetic algorithm runs were carried out using a population size of 150 and each set to carry out 25 million energy evaluations. The conformations were then clustered using a 2 Å root-mean-square fit and ranked by binding energy. From this result several structures were extracted to be used in further simulation. The first structure was the lowest energy conformer from the largest, lowest energy cluster (often considered the 'best' structure) named Model 1 (or

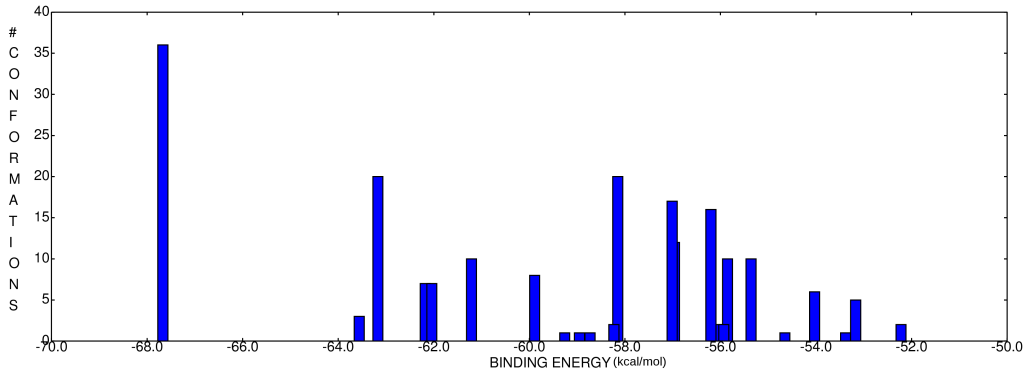

Figure 5.4: Clustering results of EGCg to cCTnC hydrophobic cleft using a refined, restricted grid and active residues chosen from NOE data.

M_1), a conformer from the same cluster but in which the ligand was rotated anti-clockwise (with respect to the aromatic ring positions in the cleft) from the position of Model 1 named Model 2 (or M_2) and a third structure named Model 3 (or M_3) from the second lowest energy cluster which was categorized by binding to the surface of the hydrophobic cleft instead of bound within it. The largest cluster is at the lowest binding energy position and is a good indicator of the success of the docking as the structure must be easily accessible and stable in order for it to be considered a likely candidate for the real docking location. A low instance of clustering would suggest that no overall position is favourable which would prevent meaningful further computational analysis. Despite the large clustering at low energy this docking was unable to reproduce exactly the experimental structure, Models 1 and 2 were both positioned slightly to the right of this and Model 3 is at the edge of the cleft. It is therefore necessary to use classical Molecular Dynamics in order to produce statistical averages which may be used to compare these models against experiment.

These were chosen as a compromise between the main factors which describe a well docked structure, namely the energy of conformation (an indication of the most 'ideal' position) and cluster size (an indication of the likelihood of the ligand reaching this position). The binding energies of Model 1, 2 and 3 were -67.67, -62.44 and -63.17 kcal/mol respectively. The docked structures are displayed within cCTnC in Figure 5.5 as well as the

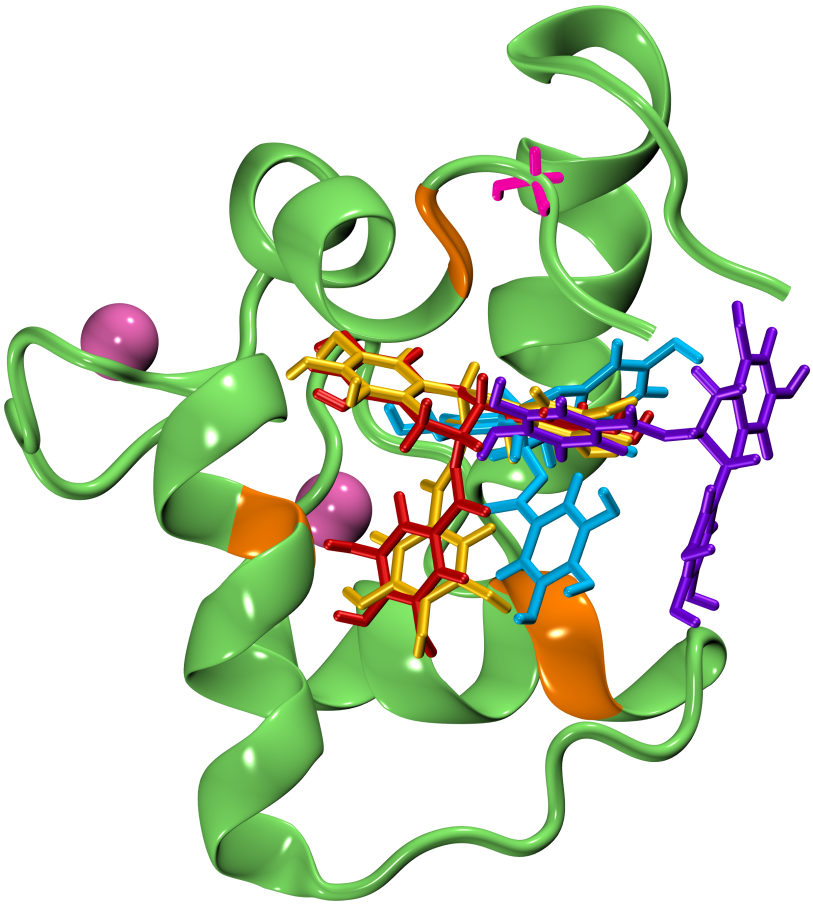

Figure 5.5: Side view of cCTnC with computationally and experimentally docked EGCg ligands. In cyan is the experimental NMR lowest energy structure; in red is M_1; in yellow is M_2 and in purple is M_3. GLY159, which is mutated to an ASP in the mutant models, is displayed in pink while the five experimentally derived NOE residues are in orange.

location of the GLY159 residue whose mutation to an ASP has been linked to some forms of Dilated Cardiomyopathy (DCM).

These conformers were chosen in order to best represent and explore the possible interactions and poses available to the catechin. Model 2 was chosen specifically due to its shifted orientation with respect to Model 1 which may change its possible interactions with the protein. Model 3 was similarly chosen to represent the catechin binding in a sub-optimal location near but not within the cleft. This may give an indication of the exit or entry pathway for EGCg in the cleft during MD.

These three models were then prepared for further classical simulation via molecular dynamics. For comparison, the cCTnC structure initially extracted from the 2KDH pdb was prepared without EGCg (the wild-type model) and is denoted as M_0. The lowest energy bound EGCg structure from the experimental NMR data was also extracted from the 2KDH pdb. This model was used to compare against the computationally docked systems and is labelled M_{exp} in the later work.

Both experimental structures were also mutated to the GLY159ASP (G159D) case in order to compare the effects the cardiac response to the addition of EGCg in this case. These models are labelled M_{0mut} (mutated troponin C alone) and M_{1mut} (mutated troponin C with bound EGCg) later work and are identical to the M_0 and M_{exp} models with the exception that residue glycine 159 has been mutated into an aspartic acid residue.

The troponin C mutation G159D was performed by taking the initial cCTnC structure and stripping all non-backbone atoms from the Glycine 159 residue and re-labelling the residue ASP. This structure was then run through *xleap*, within the AmberTools package, which is able to access the Amber database of nucleic acids and place Aspartic acid non-backbone atoms in the optimal positions around the residue (the database is composed of amino acid structure definitions based on NMR and crystallographic data as well as quantum mechanical calculations in some cases). Glycine is one of the simplest amino acids, with only a hydrogen side-chain, while Aspartic acid has a large side-chain, so care was taken to ensure that the new side-chain was not positioned within another residues inside the protein. This problem was made slightly simpler with the cCTnC model as residue 159 is very close to one of the terminal strands of the protein and so was not within a closely packed protein space. The new mutant residue was minimised within the *xleap* program to remove any close contacts with nearby atoms, then the mutated protein underwent full minimisation prior to the equilibration simulations.

5.2.1 Coordination and Inclusion of Calcium Ions

The EF-hand loop motif found in troponin C is a common feature of many signalling proteins which bind Mg^{2+} or Ca^{2+}. As previously stated in the

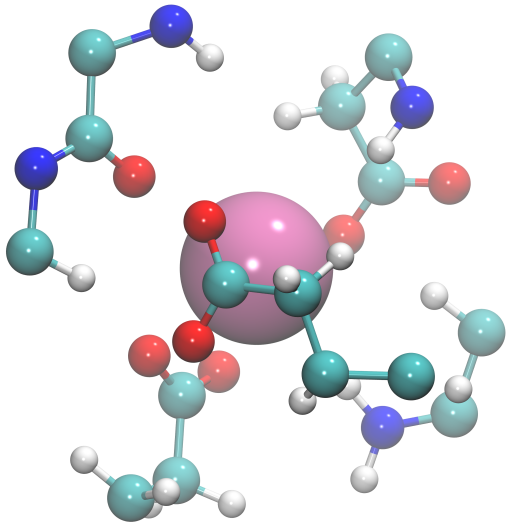

Figure 5.6: Initial coordination of the site 1 calcium (pink) with surrounding protein oxygens (red) in cCTnC. Structure taken from pdb 2KDH.

review of the troponin protein, the structural form of C-terminal troponin C contains two calcium sites which are of high affinity. In general calcium ions bind with varied coordination numbers based on the specific site; however the coordination number tends to be between 7 and 9 in protein crystallographic structures with oxygen as the preferred ligand.[57, 59, 60, 173] This is quite different compared to Mg^{2+} (with coordination number 6), which has the same charge and is known to bind to similar sites as calcium.[57, 60]

5.2.1.1 Force Field Parameters

The parameters for the two divalent calcium ions were made into a library file and remained fixed for every simulation. The Van der Waals parameter was taken from the AMBER parm99 database (originally adapted from the work by Johan Åqvist[174]) giving a radius of 1.71 Å, 6-12 potential well depth of 0.4598 kcal/mol and the charge for each atom set to 2.0. These ions were not covalently bonded to the troponin protein but allowed to move freely during the simulation. Divalent ions are difficult for molecular mechanical methods to model properly and so to test the suitability, the coordination number was monitored throughout the simulations.

5.3 Molecular Dynamics Simulations

5.3.1 Molecular Dynamics Protocols

The models in Section 5.2 were prepared for MD simulations using the program *xleap* which is included in the AMBER suite of software and incorporates many functions of this program for structure preparation and revision. The non-polar hydrogens which were removed during the docking preparation were added at minima positions based on the relevent amino acids in the protein. 11 Na^+ counter-ions were placed around the structure in order to neutralise the system (12 counter-ions were required for M_{0mut} and M_{1mut} due to the additional charge on the protein from the aspartic acid) using a Coulombic potential on a grid, followed by explicit solvation in a periodically repeated 12 Å buffer truncated octahedral box of TIP3P water molecules.[140] MD was carried out using the AMBER10[139] molecular dynamics package using the ff03[94, 105] force field. These structures were first minimised and then equilibrated followed by molecular dynamics for productions runs. The program PMEMD (or Particle Mesh Ewald Molecular Dynamics) was used to carry out all the simulations; this program is a variation of AMBER's *sander* MD engine which was designed to improve the performance of commonly used methods of simulation in AMBER (including PME simulations) as well as providing a more efficient performance on parallel machines. The simulations were run on HECToR - High-End Computing Terascale Resource, the UK's National supercomputing service.

Initial minimisation was carried out using positional restraints on the protein and ligand atoms while the water and counter-ions were allowed to relax. 20,000 minimisation steps were carried out, the first 15,000 using steepest descent minimisation followed by 5,000 conjugate gradient minimisation steps. Both speed and accuracy were considerations here as the steepest descent method is very useful for minimising potential energy of the structure quickly, but is less useful for finding the final minimum position, which the second method is much better at. Constant volume and periodic boundary conditions were used along with a cutoff of 10 Å for the non-bonded interactions. A second stage was carried out to minimise the entire system and so the positional restraints on the protein and ligand

atoms were removed. Again 20,000 steps were carried out using the same methodologies.

Following this a restrained MD simulation was carried out in order to heat the system from 0 K to 300 K using the Langevin thermostat[99] (with a collision frequency of 1.0 ps^{-1}) and at constant volume for 20 ps. Once this heating was complete the restraints were turned off and the simulation was equilibrated at 300 K for 100 ps. The SHAKE algorithm was also used to constrain the bonds with hydrogen in the protein allowing a time-step of 2 fs.[92]

Finally production runs of 50 ns of MD was carried out on each of the structures in the NPT ensemble with the constant temperature of 300 K (again using the Langevin thermostat) and constant pressure of 1 atm, using the Berendsen barostat with a time constant of 2 ps.[97, 139] Statistics from the simulations are collected for the last 45 ns of these simulations, with the first 5 ns considered an equilibration phase. In order to make selective analysis of the calcium sensitising effects of EGCg bound to cCTnC, a simulation using EMD 57033 bound deep into the cleft of cCTnC was also prepared and carried out in a manner identical to the methodology described. The structure was taken from the pdb database, code 1IH0[5], and can be seen alongside EGCg in Figure 5.7. Some discussion is given to its interaction capabilities and any altered effects on calcium ion sensitivity, including calcium coordination, in Subsection 5.3.6.

5.3.2 Secondary Structure Stability

Calculation of the RMSD of a molecular structure can provide a relatively simple and effective gauge for the stability of the molecular system as well as indicate any points of large conformational change during the trajectory. It may also be instructive to compare the protein stability in cases with and without the bound ligand in order to see the effect EGCg can have on the native state of the protein. This also allows a relatively simple comparison of the wild-type and mutated proteins. The protein backbone RMSD (atoms C, CA, N, O for each amino acid residue) for each model tested is averaged over the equilibrated 45 ns and the values, along with fluctuations, are shown in Table 5.1 and in Figure 5.8.

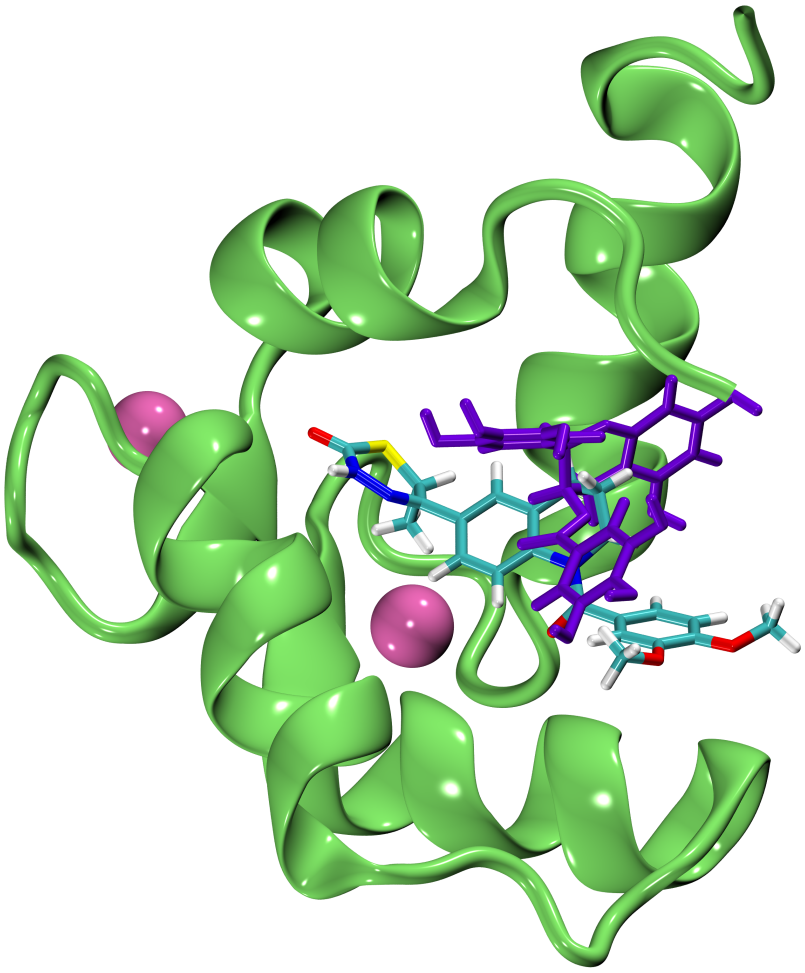

Figure 5.7: cCTnC with deeply bound EMD 57033 from the lowest energy NMR structure of pdb code 1IH0[5]. For comparison the initial M_{exp} EGCg structure is also shown in purple illustrating the comparative surface binding of EGCg.

The majority of the fluctuations appear to derive from the terminal strands of cCTnC as the centre of the protein tends to be highly stable.

Several trends can be seen from this data. It is apparent that the mutation causes both an increase in absolute deviation and overall fluctuations compared to M_0. All the models tested were found to be stable, as can be clearly seen in Figure 5.8 however the ranking of stability for the models appear to be M_{exp} followed closely by M_0 and M_1, then the M_2 and M_3 and finally the mutated systems. This is largely in accordance with what

Table 5.1: Averaged protein backbone RMSD values for cCTnC models with and without EGCg.

Model	RMSD (Å)	Standard Deviation (Å)
M_0	3.09	0.32
M_{0mut}	4.11	0.62
M_{exp}	2.89	0.27
M_1	3.09	0.26
M_2	3.57	0.37
M_3	3.93	0.40
M_{1mut}	3.92	0.40

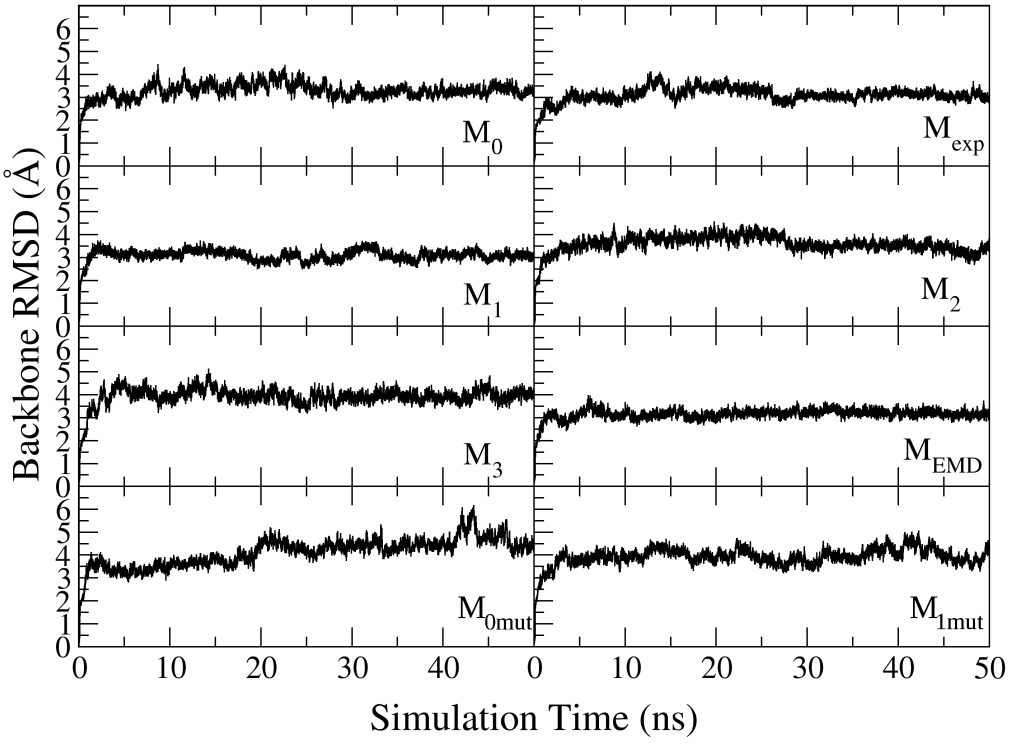

Figure 5.8: Protein backbone RMSD fluctuations over full simulation time for cCTnC models.

Table 5.2: Average EGCg hydrogen bonds formed during simulation.

Model	Solvent H-bonds	Protein H-bonds	Total H-bonds
M_{exp}	12.5	3.9	16.4
M_1	10.7	5.2	15.9
M_2	12.5	4.5	17.0
M_3	11.3	4.2	15.5
M_{1mut}	11.9	3.9	15.8
M_{EMD}	4.8	0.1	4.9

would be expected given the earlier docking results. It is also interesting to note that in all cases the addition of bound EGCg moderately reduces the fluctuations of the cCTnC protein and this is especially so in the mutated system indicating a stabilising effect of the green tea molecule on the protein structure.

5.3.3 Hydrogen Bonding

As EGCg contains 8 hydroxyl groups and 3 further oxygens it is expected that hydrogen bonding will be a major contributor to the interactions between it and the local protein environment. Indeed, in Chapter 3, we find that in classical water solvent EGCg forms almost 17 inter-molecular H-bonds in total along with a small amount of intra-molecular bonding as well. When bound to cCTnC, the EGCg rings appear highly conformationally restricted in a hydrophobic environment so we compare these differences. In addition to the standard hydrogen bonding atoms (O and N in the case of cCTnC) it has also been shown that this type of bonding can occur with sulfur atoms.[148] As cCTnC contains sulfur atoms in the Methionine sidechain, one residue of which has been shown experimentally to have NOE contacts with EGCg, we include the possiblity of hydrogen bonding between the two. EMD 57033 contains four oxygens, three nitrogens and a sulfur atom however these are generally restricted in ring positions or in the case of the oxygens, have attached methyl groups. This prevents excess interactions by hydrogen bonding and, as was seen in Chapter 3, fewer interactions compared to EGCg.

The total H-bonding to the EGCg ligand when in complex with the

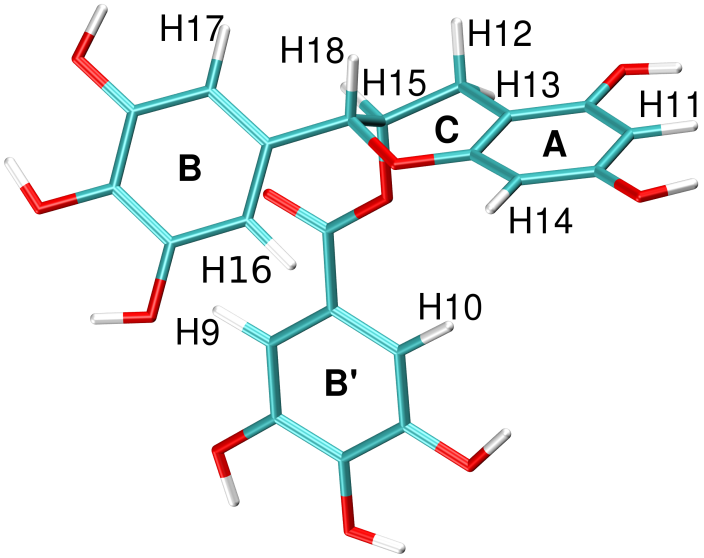

Figure 5.9: EGCg molecule with labelled rings and carbon bonded hydrogen as indexed in AMBER.

cCTnC hydrophobic cleft is slightly reduced compared to the solvation alone in water except in the case of M_2 where it is almost identical. The larger number of solvent H-bonds of M_2 and M_3 are likely due to a slightly more loosely bound position on the protein when compared to M_1 as established from the docking simulations. The M_2 results closely agree with the M_{exp} structure and it is highly probable that the M_2 EGCg is exposed to the solvent in the same manner due to their similar starting ring orientation with respect to the cleft. M_1 shows the largest number of H-bonds with the protein, localised primarily on the A and B' rings (the rings are shown labelled in Figure 5.9). It appears that these strong H-bonds prevent a re-orientation of the rings which may act to reduce overall ligand exposure to solvent. The hydroxyl groups on EGCg will readily donate to form an H-bond throughout the entire simulation except in the case of M_1 where it is likely the tight binding of the A and B' rings trap one of the OH groups inside the hydrophobic cleft. These results suggest that orientation of the ligand rings with respect to the protein has a much larger influence on total H-bonding than tight binding to the protein via initial docking.

As expected, EMD 57033 did not interact with cCTnC through hydrogen

Table 5.3: Major protein-ligand H-bonds formed during simulation.

Model	Major H-bonded Residues
M_{exp}	GLU161, GLY159, MET120
M_1	PHE156, GLY159, GLU161
M_2	GLU126, GLU161, GLY159
M_3	GLU135, GLY140, ASP139, THR124
M_{1mut}	GLY91, MET120, LYS158

bonding in any significant way. The ends of the molecule which are exposed to solvent are able to form hydrogen bonds and it is here that the majority of the H-bonds are found, with the ring bound nitrogens and sulfur adding only around 0.3 H-bonds to the total. Overall H-bonding in the M_{1mut} case is very similar to M_{exp} as we expect but it is likely that the conformational re-organisation of the protein due to the mutation has reduced the possible solvent exposure to EGCg compared to M_{exp}.

Table 5.3 discloses the important protein-ligand H-bonds which contribute to the stability of the system. These were defined as H-bonds which existed for more than 30% of the total simulation time. For the unmutated models we generally see the same two residues reappearing, GLY159 and GLU161 on the terminal H helix, though as expected this is not the case for M_3 which attaches more closely to the G and F helices of cCTnC. These interactions with the terminal strand of the protein are likely damping the fluctuations of this part of the protein and may partially explain why the M_3 RMSD was much higher than the others. The M_{1mut} case is slightly different in that the GLY159 is now ASP159. Due to the larger size of the ASP side-chain it appears that it becomes more preferrable for EGCg to interact with the adjacent residue LYS158 instead. In addition it appears that this allows EGCg to form an H-bond with GLY91 at the terminal near helix E, meaning EGCg spans both sides of the cleft while still interacting deeply within the cleft to MET120. This then leads to the reduction in fluctuations we see in Table 5.1 for the mutated case.

Table 5.4: $\pi - \pi$ Interactions during production runs between EGCg and cCTnC PHE residues.

Model	Ligand Ring	PHE Residue	Bond lifetime
M_{exp}	A	PHE104	0.1%
M_1	A	PHE156	0.6%
	B	PHE104	0.3%
M_2	B′	PHE156	4.8%
M_3	A	PHE156	43.1%
	B′	PHE156	67.8%
M_{1mut}	B′	PHE156	1.1%
M_{EMD}	A	PHE156	32.0%
	A	PHE104	4.3%
	D	PHE156	81.5%

5.3.4 Aromatic Stacking Interactions

The cCTnC protein contains two Tyrosine and four Phenylalanine residues which may undergo aromatic stacking with EGCg or EMD 57033. While a π stacking interaction is diffcult to accurately describe using classical MD (due to the lack of correlation effects included in the model) it is still possible to observe analogous interactions from the van der Waals interaction between aromatic rings. These rings were tracked throughout simulation and their relative distances and angle to the ligand rings monitored. When two aromatic rings came within 5 Å (measured centre-to-centre) of each other with the angle between the vectors normal to those rings less than 45°, a stacking interaction was considered to be taking place.[175] The only aromatic rings that showed interactions with the EGCg ligand were those of the PHE104 and PHE156 residues which remain close to the surface of the hydrophobic cleft. These can be seen with respect to the EGCg binding site in Figure 5.10.

Table 5.4 gives a full breakdown of the aromatic stacking interations taking place in the models. PHE156 appears much more likely to interact with the ligand and in one case did so for a significant percentage of the total simulation time. The majority of interactions taking place in this manner were nearly instantaneous where two rings come briefly into contact but do not maintain the interaction for very long. Despite this, stacking may

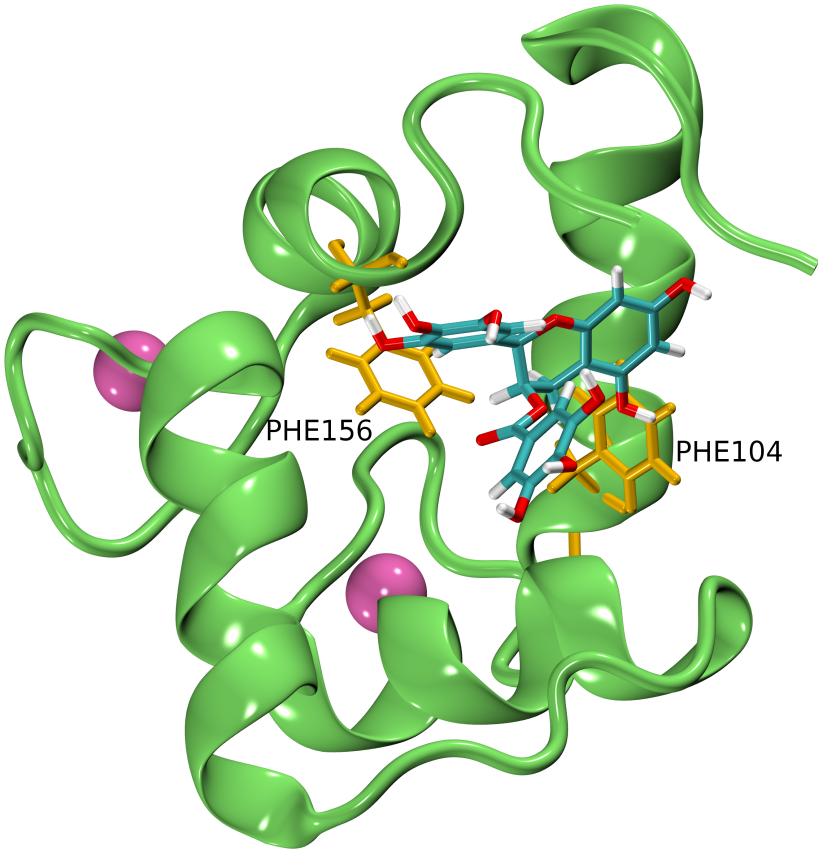

Figure 5.10: cCTnC with labelled aromatic residues which lie on the hydrophobic cleft surface. M_{exp} is also shown for comparison.

be significant as there are some changes in the protein arrangement when stacking occurs due to the space required for the rings to rotate sufficiently to form a bond in the first place.

This is especially noticable in the case of M_3 where the loose binding of the ligand allows the rings to easily re-arrange into favourable stacking positions. The lower number of protein-ligand H-bonds may also have affected the ability of M_3 to rotate to accomodate such stacking interactions. Notably, in the M_1 case, a strong H-bond is seen to form between PHE156 and EGCg which would prevent the rotation required to stack the rings. Given the wide variance in the frequency of aromatic stacking occurring between the EGCg and cCTnC for all models it is unlikely that the interaction plays a dominant role in the binding process between the two in this case

though these interactions may help compensate for sub-optimal binding. In the case for M_{EMD}, when monitoring the two EMD 57033 aromatic rings (B and D in Fig. 3.21) as well as the sulphur-containing ring (A), the average number of $\pi - \pi$ interactions was 1.2; these were essentially with PHE156, and to a much smaller degree with PHE104. The increased amount of $\pi - \pi$ interactions compensated for fewer viable hydrogen bonds with the protein with respect to EGCg.

5.3.5 Comparisons with Experimental NOE Data

NMR imaging is an experimental technique which is very useful for generating structural conformations of large proteins and other polymers. Using the Nuclear Overhauser Enhancement (NOE) distance restraints (generated from NOESY data) as a comparison, the computationally docked simulations can be assessed for their suitability to represent the EGCg-troponin complex. The NOE effect is proportional to $\frac{1}{r^6}$ where r is the inter-atomic distance. Therefore by tracking the inter-atomic distances between atoms during the simulation a set of distance values comparable to those inferred from experiment can be determined which should estimate the average position of the ligand with respect to the cCTnC. These are shown in Figure 5.11. The naming convention for the EGCg H atoms is shown in Figure 5.9.

Due to the symmetry of the EGCg molecule several hydrogen atoms are considered indistinguishable in experimental NOE calculations such as the H16 and H17 as well as the H9 and H10. This is not a problem in computational studies as the atom indexes are unique, however in order to produce a representative set of data which is comparable to experimental work the r^{-6} data have been averaged together where appropriate.

As we expect, due to the stochastic nature of the MD, the simulations do not match exactly with the NMR data; however it should be noted that in order to show a true agreement between experiment and simulation the simulated NOE distances should not only reproduce those NOEs found by experiment but should also not show any NOEs that are not also found in experiment.[176]

The M_{exp}, M_1 and M_2 results appear to be in reasonable agreement with the NMR data. The M_{exp} structure is in closest agreement with this (as

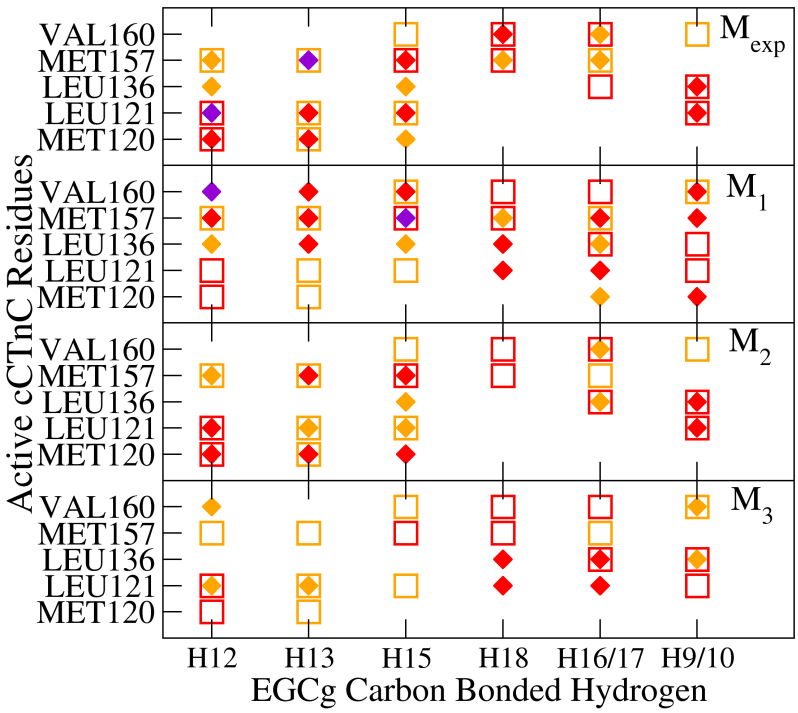

Figure 5.11: Comparison between experimental NOE data (empty squares)[6] and calculated (diamonds) NOE distances between select hydrogen atoms of EGCg (abscissa) and cCTnC residues (ordinate) in the cCTnC·EGCg models. The colours correspond to the following ranges: medium (2.7 Å - 3.3 Å, purple); weak (3.3 Å - 5.0 Å, red); very weak (5.0 Å - 6.0 Å, orange).

we might expect) as seen in the similarity of the H12 and H13 contacts to MET120, LEU121 and MET157 as well as the H18 and H9/10 NOEs. The M_1 results appear slightly shifted away from the M_{exp} structure, likely due to the different polyphenol ring orientation with respect to the cleft as discussed in Section 5.2, suggesting that despite different docked orientations the average position of EGCg during these simulations is not markedly different. There are striking similarities between the M_{exp} and M_2 NOEs, notably with the H12/H13 contacts to MET120 and LEU121 indicating that both structures during simulation do not move significantly far from their initial position and orientation, but it is likely that the M_2 structure moves to a slightly more central position in the cleft from its initial one in order

to reproduce these NOE results so closely.

The M_3 trajectory, on the other hand, clearly diverges with the others as shown by a complete lack of interaction with MET120 and MET157, again likely due to its original docked position being further out of the hydrophobic cleft. This indicates that the large potential barriers to rotation of the polyphenol rings examined in Chapter 3 do not prevent the structure from finding an equilibrium position in the cleft, nor does binding to different parts of the troponin C hydrophobic cleft prevent the EGCg from optimally orienting within the simulation time once there. These results give credence to the notion that the EGCg position is stable in the hydrophobic cleft and does not require a specific ring orientation with respect to the protein, though molecular interactions can only slightly adjust the EGCg position in the cleft and will not completely compensate for poor binding to the protein.

5.3.6 Calcium Sensitivity

5.3.6.1 Calcium Fluctuations

Calcium ion sensitivity is at the core of questions over effective treatments for heart disease which target the troponin complex, whether by direct effects on the binding propensity of inter-cellular calcium ions or by less direct modifications of the protein-protein interaction causing changes to the calcium signalling response. While it is known that the calcium sites in cCTnC are of high affinity and do not instigate the contraction-relaxation cycle in the same manner as the calcium binding to the N terminal domain, it has been shown that mutations which affect calcium sensitivity on the N terminal will also affect the C terminal sites and thus the reverse is also likely to hold true.[177]

The behaviours of the calcium ions at the two binding sites, which we labelled I and II in Figure 5.3, were monitored and compared for the various models. The distributions of the RMSD of each calcium ion calculated with respect to their positions at the end of the MD equilibration for all the models aligned with respect to the cCTnC backbone are shown in Figure 5.12. For the M_0 simulation, deformation to reduce solvent exposure of the

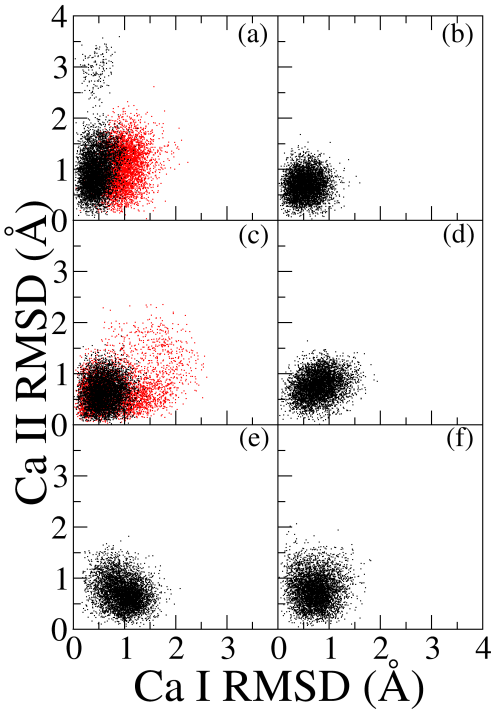

Figure 5.12: Distributions of the RMSD of calcium ion at site I versus that of site II. Overlaid in black are the wild-type cCTnC simulations while in red are the mutated troponin simulations: (a) M_0, (b) M_{EMD}, (c) M_{exp}, (d) M_1, (e) M_2, (f) M_3.

hydrophobic core tended to confine the calcium ion at site I while having the opposite effect on site II. The destabilising effect on site II was so large that the corresponding calcium ion began a transition to a new location after nearly 48 ns (the simulation was extended a further 4 ns in order to map this transition). For comparison the results of the EMD 57033 simulation are shown where it can be seen that the presence of the ligand highly restricts fluctuations of the ions. The addition of EGCg to the hydrophobic cleft tended to drastically reduce fluctuations on both calcium sites, including the case of binding in M_3, although it is further from the cleft than the other models. Interestingly, the G159D systems with and without EGCg showed larger fluctuations overall, especially for the calcium ion at site I, though no transition to a different location was observed. This is consistent with the experimental observation that calcium ions show a reduction of

Table 5.5: The coordination numbers of calciums ions averaged throughout the cCTnC MD simulations.

Model	Calcium 1	Calcium 2
M_0	8.7	8.8
M_{0mut}	8.3	8.8
M_{exp}	8.7	8.9
M_1	8.5	8.9
M_2	8.8	8.8
M_3	8.5	9.0
M_{mut}	8.7	9.0

binding affinities in cardiac tissue affected by dilated cardiomyopathy.

5.3.6.2 Calcium Coordination

The number of oxygen atoms coordinated to the calcium ions was calculated by integrating the radial distribution function of calcium to oxygen up to the first minima which was found to be a distance of 3.5 Å. The shell peaks were seen at 2.65 Å. Calcium coordination distance is slightly different depending on the protein/environment and this distance appears to be in reasonable agreement with previous work on the coordination distance of calcium ions in water under various geometries as well as coordination in EF hand loops.[178, 179] The results for the cCTnC models in Table 5.5.

There appears to be very little variation in the coordination between simulations, including those of the mutated protein. The coordination of the site 2 calcium does not change by more than 0.2 for all the simulations tested, while for site 1 this is slightly lower, with a wider variation of 0.5. For both sites the values are within the range of 7-9 expected from crystallographic data and so it appears that the simple model used to treat the divalent cations is sufficient.

5.3.7 Binding Free Energies

The binding free energy of the protein-ligand interaction was evaluated using the MM/PBSA facility within the AMBER program suite.[180] This program combines molecular mechanics energies with continuum solvation

Table 5.6: Estimates for the free energy of binding between cCTnC and EGCg using an internal dielectric of 1.0. All values given in kcal/mol.

Model	ΔH	$-T\Delta S$	ΔG	$\Delta\Delta G$
M_{exp}	-16.3 ± 5.8	20.2 ± 4.3	3.9 ± 7.3	0.0
M_1	-18.6 ± 5.8	19.7 ± 4.2	1.0 ± 7.2	-2.9
M_2	-16.8 ± 6.6	19.1 ± 3.0	2.3 ± 7.3	-1.6
M_3	-6.5 ± 5.8	18.9 ± 5.0	12.4 ± 7.7	8.5
M_{1mut}	-16.4 ± 4.3	19.0 ± 3.8	2.4 ± 5.7	-0.1
M_{EMD}	-22.3 ± 4.1	20.1 ± 2.5	-2.2 ± 4.8	-6.1

model approaches (discussed in Chapter 2) within the Poisson-Boltzmann Surface Area (PBSA) scheme. A salt concentration of 0.1mM was mimicked in order to more closely reproduce the biological environment. Enthalpic contributions were evaluated using the PBSA method while estimates of the entropy at 300 K were calculated with normal mode analysis. Snapshots for both calculations were extracted from the production runs in steps of 10 ps for the enthalpic term and 2.25 ns for normal mode analysis. These were chosen to remove correlation effects in the enthalpic portion and as a compromise on the large calculation time for the entropic portion whilst attempting to maintain a good degree of accuracy. There are a number of drawbacks to the MM/PBSA method which are thoroughly discussed in Chapter 2 but the method has shown to provide an effective estimator for ranking binding of similar structures without consuming exorbitant computational resources, especially useful in cases where there are proteins in solvent.[181, 182] Calculations were performed using an external dielectric of 80 to represent water solvent and an internal dielectric of 1.0 and 2.0 to represent the complex. The default value of 1.0 for the internal dielectric is commonly used for proteins and polymers to simulate their generally hydrophobic interior, however it has been shown that the accuracy of enthalpic calculations can be improved by increasing this in cases where there are a large number of charged residues.[119, 183, 184] For cCTnC 43% of the residues are explicitly charged, with a considerable number of polar residues in addition to this and so calculations with both dielectrics were carried out to test their effectiveness. The results are displayed in Tables 5.6 and 5.7.

The final column in each table gives a relative ranking using the M_{exp} as the baseline to allow a quick estimation of the models binding character relative to the Experimental structure. Entropy values for all the systems are very similar as we might expect for identical molecules and the main difference here lies in the vibrational component of the complex, indicating that EGCg movement becomes more or less restricted in the cleft depending on the docking. The overwhelming difference in binding energy is therefore derived from the enthaplic contribution. These values are quite similar for all models except M_3 for which it is drastically different, likely due to the change in local protein environment as well as a larger prevalence for π stacking interactions during the simulation. It is interesting to note that the computationally docked M_1 model indicates a more favourable binding than the structure proposed experimentally while M_{exp}, M_2 and M_{1mut}, which have very similar docked positions, are effectively equivalent. This indicates that the rotated bound structure of M_1 might be a better structure for the bound state of the system overall. In addition it should be noted that the associated deviations in these values is quite high, in fact generally larger than for the entropic contribution (this is generally the larger source of error in these type of calculations). Moreover, in terms of absolute ΔG Table 5.6 suggests that the binding is unfavourable and EGCg will dissociate quickly, especially for M_3 which we might expect to un-bind within the simulation time. This is of course not seen for any models and experimental binding to cCTnC ascribed a K_d (dissociation constant) of 1.1mM which when converting into a free energy of binding by $\Delta G = RTlnK_d$, translates into a ΔG value of -4.06 kcal/mol indicating a moderately favourable bound state. It is therefore worth analysis of the effects of using a higher internal dielectric, shown in Table 5.7.

In this table it is immediately apparent that a more consistent view of the enthalpic terms is seen with smaller deviations from the M_{exp} baseline. We also achieve lower fluctuations and more favourable overall ΔG (though it appear now to slightly overestimate this absolute value) which is favourable within error margins (except for M_3 which is slightly disfavoured). The relative ranking of the binding is fairly similar though M_2 is no longer more favourable than the M_{exp} system. The overall difference between M_1 and

Table 5.7: Estimates for the free energy of binding between cCTnC and EGCg using an internal dielectric of 2.0. All values given in kcal/mol.

Model	ΔH	$-T\Delta S$	ΔG	$\Delta\Delta G$
M_{exp}	-27.6 ± 5.2	20.2 ± 4.3	-7.4 ± 6.7	0.0
M_1	-27.7 ± 4.5	19.7 ± 4.2	-8.1 ± 6.2	-0.7
M_2	-25.0 ± 3.6	19.1 ± 3.0	-5.9 ± 4.7	1.5
M_3	-21.9 ± 3.5	18.9 ± 5.0	-3.1 ± 6.1	4.3
M_{1mut}	-28.1 ± 3.5	18.8 ± 3.8	-9.4 ± 5.1	-2.0
M_{EMD}	-31.1 ± 3.7	20.1 ± 2.5	-11.0 ± 4.5	-3.6

M_{exp} is now smaller and largely dependant on the change in entropic value due to the alternate ring orientations. In both dielectric calculations we find that the G159D mutation increases the binding propensity for EGCg over the wild-type situation. We also confirm that deep binding to the cCTnC cleft is more favourable than binding to the surface, as suggested by experimental and docking studies. Due to the much smaller discrepancy between the experimental and M_{exp} ΔG, as well as a moderate improvement in standard deviation of all the results, it is apparent that an internal dielectric of 2.0 provides a better treatment of the cCTnC·EGCg molecular system.

5.4 Studying the Conformational Landscape of EGCg in the Hydrophobic Cleft using Metadynamics

5.4.1 Dihedral Collective Variables

Well-tempered metadynamics calculations were carried out in order to explore the Free Energy Surface (FES) of the protein ligand system as well as explore alternative orientations of the aromatic rings of EGCg given their variability in docking outcome. The well-tempered scheme was chosen due to its efficient sampling speed and reduced error in FES reconstruction.[109] This was undertaken using the PLUMED plug-in for AMBER10.[108] In preparation for this several tests were carried out in order to determine the most appropriate set of parameters for the simulations and it was seen

that the use of a 2 fs time step with SHAKE on did not noticeably affect the mapping of the free energy surface but did increase the conformational space searched without a corresponding increase in computational time required. Previous dihedral analysis has established that the position of the polyphenol on troponin C is dependent upon three torsions which describe the positions of the phenolic rings respective to one another. These were therefore used as the collective variables (CVs) to be specified in the metadynamics simulation. The well tempered run was carried out at a simulated temperature of 300 K; they included a bias factor of 15 kcal/mol with a Gaussian height of 0.1 kcal/mol and width of 0.35 radians. The Gaussian deposition τ_G was set to be 500 time-steps and in all other respects the molecular dynamics was carried out as previously using a Langevin thermostat and Berendsen barostat. The starting configuration was chosen to be the M_{exp} after 30 ns of simulation. It is not strictly necessary to take the structure after such a long MD as the system need only be thermally well equilibrated, however in order to ensure the protein and complete complex were fully relaxed before beginning metadynamics extended time was given after equilibration before extracting the initial structure.

Previous metadynamics work to elucidate the free energy surface of EGCg in vacuo and water can be seen in Chapter 3. Free energy profiles of the EGCg in the cCTnC hydrophobic cleft as a function of the torsional CVs can be seen in Figure 5.13 after a total of 120,000 Gaussians were deposited. The profiles shown are averaged over the last 1,000 profiles in order to reduce the associated error to fractions of kcal/mol which, when considered with the effect of the well-tempered scheme to reduce gaussian height as a function of bias factor when in an unchanging collective variable space, gives an accurate estimate for the free energy surface. As previously, the 1-D profiles were calculated according to Equation 3.2.

Also shown are the reconstructed free energy surfaces averaged over 1,000 profiles in order to further reduce the error in the free energy values. It appears that the combination of solvent and protein interactions cause the CV minima and maxima to shift slightly away from the 0°, 90° and 180° clearly shown by the θ and ϕ CVs in comparison to the EGCg alone in vacuo and solvent. It is also noted that the θ and γ potential barriers are

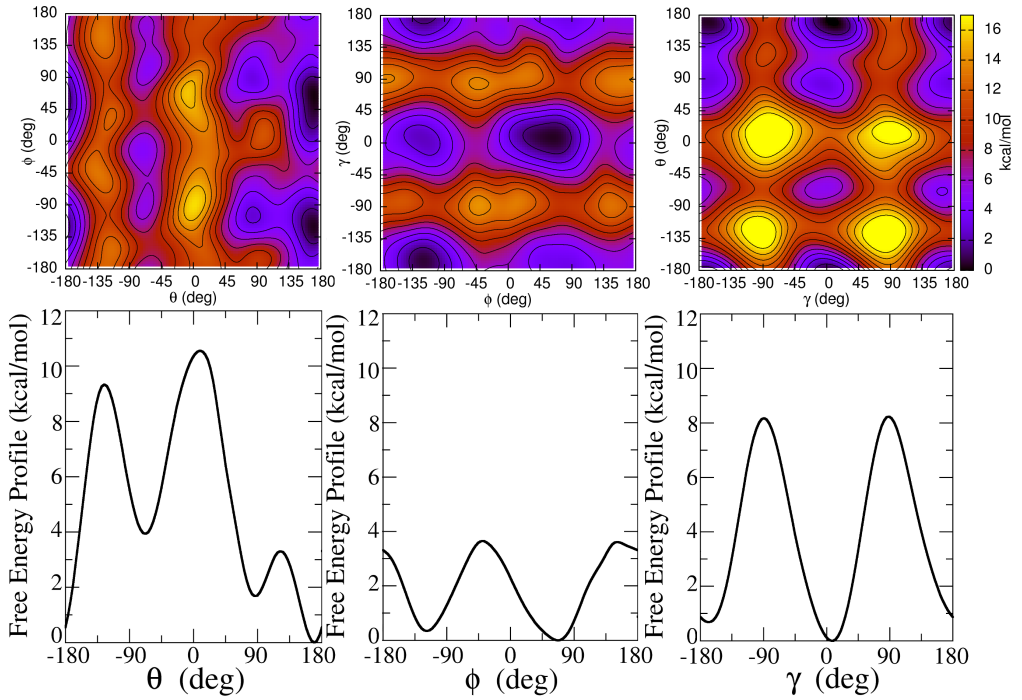

Figure 5.13: Top: Free energy surface as a function of the dihedral angles θ, ϕ, γ with the energy scale set to the absolute barrier minimum and contour spacing of 2 kcal/mol. Bottom: Energy profiles of the dihedral torsions (from left to right: θ, ϕ, γ) as collective variables projected over time and setting their absolute minimum to zero.

too high (total current barrier height is 10.6 and 8.2 kcal/mol respectively) for them to be properly explored using conventional MD and this explains why these angles were not seen to change during the MD (except θ between the smaller 180° and 90° barrier). The angle of the ring B′ can therefore be considered relatively fixed after docking to the protein. The θ profile has changed considerably, the maxima at -135° is 3 kcal/mol higher while the 0° maxima has decreased in height by nearly 2 kcal/mol. This, combined with the angular shifts, suggests that inter-molecular interactions, both of the solvent water and protein, have a more dominant effect on the dynamics of this ring. The ϕ angle however shows a much lower barrier of 3.7 kcal/mol which in terms of barrier height is in reasonable agreement with the barrier profile obtained during the earlier metadynamics of EGCg alone. The slight shifting of the maxima points is again likely due to the close proximity of the

Table 5.8: The standard deviation associated with the solvated cCTnC·EGCg torsional metadynamics simulation (kcal/mol).

Profile	Standard Deviation
EGCg γ	0.331
EGCg θ	0.227
EGCg ϕ	0.321

hydrophobic cleft of cCTnC preventing free rotation of the ring. The low barrier height will have contributed to the high angle variance in the dihedral analysis earlier. The number of H-bonds formed with solvent during this time was 12.77 in very close agreement with the MD results (the slight discrepancy may be attributed to the forced rotation of the rings breaking protein-ligand H-bonds and briefly exposing the OH groups to more of the solvent while in the cleft). Overall it is apparent that while protein-ligand interactions will have an effect on the free energy surface, these effects do not change the general shape of the profiles. The standard deviation for each profile can be seen in Table 5.8.

Table 5.8 shows fluctuations an order of magnitude larger than those in Table 3.2 which, though still small, indicates that the simulations required longer time-scales to converge. As EGCg would have be able to interact with a number of nearby protein residues, it is likely that these interactions would require more simulation time to fully explore the full energy landscape than for the isolated catechin. Despite this increase, it appears that the use of the well-tempered scheme has produced a good estimation of the profile energy barriers of EGCg bound to cCTnC with minimal error.

5.4.2 Distance Collective Variables

Metadynamics was carried out using a restricted selection of the previous NOE distances discussed in Section 5.2. This was done in order to explore the overall energetic favourability of EGCg to bind and stay within the cleft. A preliminary simulation was undertaken in which no other restraints were applied to the system and was set up as follows: the Gaussian height was 0.1 kcal/mol, width was set to be 0.5 Å and Gaussian deposition rate was

kept at 500 steps. The distance restraints were selected in a way which allowed them to represent the surface contact of the EGCg rings with the hydrophobic cleft of cCTnC. The chosen residues are shown in Figure 5.14. The metadynamics run was set up similarly to that of the dihedrals except in every case the CV was chosen to be the averaged distance between the side-chain hydrogens capable of forming NOEs and carbon bonded hydrogens attached to specific rings of EGCg:

CV1: The MET120 residue distance to the H12/H13 atoms on the C ring.

CV2: The LEU136 residue distance to the H9/H10 atoms on the B′ ring.

CV3: The VAL160 residue distance to the H16/H17 atoms on the B ring.

The preliminary CV results are shown in Figure 5.15. The 1-D profiles were calculated according to Equation 3.2. The free energy minimum position appears to occur when CV3 remains near 4.5 Å though there also exists meta-stable states of below 6 Å for both CV1 and CV2. The minima state at 3.5 Å for CV1 indicates a probable weak protein-ligand H-bond interaction affecting the A/C ring which is easily broken as shown by the very small barrier between that and the overall minima at 6 Å. The barrier for CV2 at 6 Å is larger but once passed, the ring appears free from protein interactions as indicated by the broad global minima.

From inspection of the profiles it should be possible to introduce potential walls at 12 Å distance without a large disturbance to the relevant FES profile, allowing us to increase free energy surface resolution by reducing gaussian width to 0.25 Å without extending computational time unnacceptably. This was carried out by adding artificial CV energy potential walls of 100 kcal/mol at 12 Å distance. The well-tempered scheme was carried out with a bias factor of 10, anticipating the smaller barriers between meta-stable states seen in the preliminary work. In other ways the simulation was identical to the preliminary set-up and in total 80,000 gaussians were deposited. The reconstructed free energy surfaces are averaged over 100 profiles and shown in Figure 5.16.

The reconstructed FES for this simulation contains several areas of interest. Regions in yellow were not explored at all during the metadynamics,

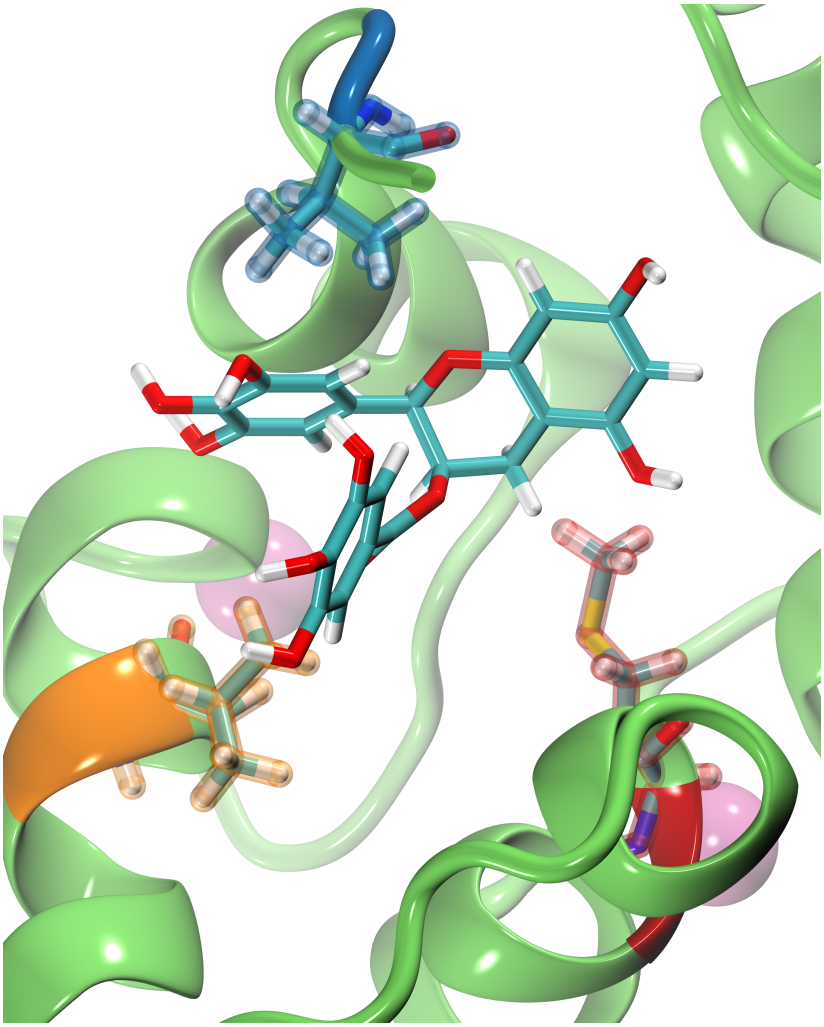

Figure 5.14: Close-up of the cCTnC cleft with bound EGCg. Arranged around are coloured representations of the protein residues used as distance collective variables. MET120 is lined in red, LEU136 is lined in orange and VAL160 is lined in blue. Close proximity of the side-chain hydrogens with the relevent carbon bonded hydrogens on EGCg can clearly be seen.

in all likelihood due to them being inaccessible as a combination of CVs (because they would involve overcoming the short-range interaction or because they would stretch EGCg into a highly energetically infeasible conformation), while red areas were only sparsely explored due to a lack of potential molecular interactions to favour such a state. There are several meta-stable regions within the 10 Å and one beyond, indicating a possible transition

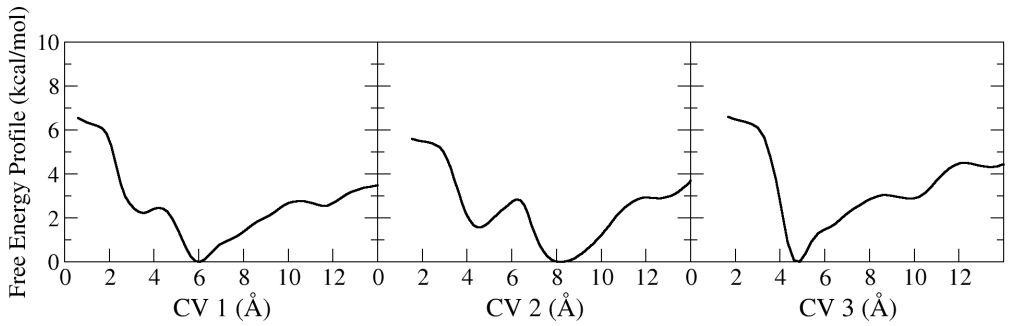

Figure 5.15: Preliminary free energy profiles of the three distance CVs used in metadynamic simulation.

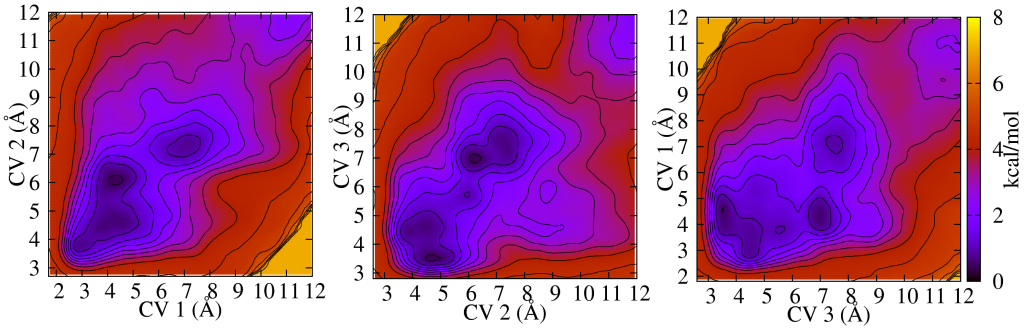

Figure 5.16: Reconstructed FES as a function of collective variables for the distance restraint metadynamics simulation with potential walls.

pathway exiting the hydrophobic cleft. Within the 10 Å region there appear to be 2 main energy minima regions which likely comprise an amalgam of several favourable protein-ligand interactions. These are at the very close 2.5 Å - 4 Å and more distant 6 Å - 8 Å regions. The very close corresponds to direct interactions between the relevent EGCg aromatic and the protein residue involved in the CV while the distant is the global minima and indicates a number of interactions with other close by protein residues. This suggests that within the hydrophobic cleft EGCg can remain moderately mobile and is likely to form several favoured conformations within it with no single one being overwhelmingly preferred. The exit pathway was not explored fully and so an estimate for the energy barrier could not be made. This is primarily due to the need to impose the potential walls to limit the

Table 5.9: The standard deviation associated with the solvated cCTnC·EGCg distance metadynamics simulation (kcal/mol).

Profile	Standard Deviation
EGCg CV1	0.212
EGCg CV2	0.604
EGCg CV3	0.253

computational cost. It may be possible in future to examine this transition by imposing walls further out, at 15 Å or, better, at 20 Å however the implied computational cost would be severe and it may instead be more useful to consider alternate methods such as a funnel metadynamics regime. An estimation of the error in the energy profiles has been performed in the same manner as described in Section 3.3.3 and the errors are displayed in Table 5.9.

The magnitude of the errors in Table 5.9 are the same as those seen in the dihedral metadynamics of EGCg bound to cCTnC, shown in Table 5.8. It is likely that this is for similar reasons, as a wide number of interactions are possible between each of the aromatic rings of EGCg. As can be seen in the 2D free energy surface in Figure 5.16, this lead to a number of small minima positions and this may explain the larger fluctuations here. In addition, the introduction of the potential walls will have increased the uncertainty in the profiles at distances close to the 12 Å wall. Overall however the error does not appear to impact the profiles to a significant degree.

5.5 Chapter Summary

Using experimentally derived work as a base it was possible to show that computational docking could reproduce a close approximation of the original binding (with M_2) as well as provide a favourable alternative (M_1) which both corresponded closely to NOESY restraints, a highly sensitive measure of structural position. MD simulations provided a wealth of detail on the potential interactions for the system such as the effects of mutation and ligand binding on calcium affinity as well as allowing a reasonable estimate of the ranked binding energy, ΔG, of the docked structures where it was

also shown that protein enviroment needs to be carefully considered when making these estimates. The simple model used to describe the calcium ions bound to the protein was also found to be successful. Metadynamics simulations indicated that while the free energy profile is shifted slightly in the protein cleft, it is not distinctly affected and the torsional barriers inherent to the EGCg molecule remain the primary driver of the conformational state and will determine the specific inter-molecular interactions which take place. This also allowed us to monitor the effect of rotation on binding propensity where it was found that torsional rotation within the protein cleft does not led to dissociation of the system, due in large part to the alternate meta-stable positions available. The distance restraint metadynamics gave an indication of a possible dissociation pathway, despite reducing the major interaction of the system to 3 collective variables, elucidating the further potential of such methods in these systems.

Chapter 6

Interaction of Epigallocatechin 3-Gallate with the Cardiac Troponin C - Troponin I Complex

6.1 Chapter Overview

In this Chapter the protocols used in Chapter 5 have been extended to the troponin C - troponin I complex, a larger section of the entire troponin protein. Computational docking techniques are used to identify an alternative binding site for the EGCg catechin in this case and MD is performed to test the effects of this binding on the calcium mobility and protein interactions. This work has been published in PLOS ONE.[164]

Sparse data from experimental NMR studies[6] are used in conjunction with the computational binding data to establish a possible second binding site as well as a double-bound protein-ligand system which is further tested with MD. These results are discussed in terms of the specificity of ligand binding as well as the collaborative abilities of experimental and computational work to refine difficult to resolve cases such as this.

6.2 The Troponin C - Troponin I Complex

The protocols used in Section 5.2 have shown to be able to correctly reproduce the experimentally verified binding of EGCg to cCTnC as well as favourable alternatives within the cCTnC hydrophobic cleft. It should be noted that this hydrophobic cleft is also where a fragment of cTnI binds to cCTnC and is responsible for passing the calcium signalling response to the extended cardiac system. With the knowledge gained in the previous Chapter it becomes useful to apply the same protocols to this extended cardiac system. A more complete picture of the cardiac troponin protein structure is available in the experimental X-ray crystallographic structure in the PDB database with ID 1J1D[185]. This is shown in the previous Chapter in Figure 5.1. In the native state the protein subunit cTnI binds cTnC to the rest of the muscle fibre. The portion of cTnI which achieves this is known as the cTnI(34-71) fragment. This structural binding exists in the hydrophobic cleft of cCTnC where EGCg is seen to bind in Chapter 5. It has been experimentally verified that known calcium sensitisers such as EMD 57033 do not bind concurrently with cTnI(34-71).[6] Attempts to experimentally discern a viable binding site for EGCg in this case were also unsuccessful[6] though several chemical shifts were observed hinting that some form of binding does take place. Thus computational simulations may provide a viable alternative given their success in the previous Chapter. In the first part of this Chapter docking and molecular dynamics of the cCTnC·TnI(34-71) with EGCg is carried out and the effects of the ligand binding as well as protein mutation are discussed with reference to their effect on calcium signalling transmission. Further comparison to calcium mobility with regards to the new binding site is also made. In the second part of this Chapter the new binding site is further developed with aid from the original experimental data and a double-bound protein-ligand structure is proposed. Using an up-to-date force field, designed for enhanced protein stabilisation, both binding sites are investigated through MD and the effects are analysed with respect to the purely computational results. Finally suitablity for this structure to represent the full experimental data is discussed.

6.2.1 Docking Protocols

A similar recipe to that found in Section 5.2 was carried out. The structure used was taken from the previously mentioned 1J1D pdb where only the cCTnC and cTnI(34-71) sub-units were retained. Again, all non-polar hydrogens were removed prior to docking and the charge scheme remained the same with default Gasteiger charges on the protein residues and DFT charges applied to the ligand. In this case, as there was a lack of clear NOESY experimental data to refine the search, an all-rigid receptor docking with full grid surrounding cCTnC was undertaken with a 10 Å grid buffer on all sides of the cCTnC protein. As before a grid spacing of 0.182 Å was used, 200 Lamarckian genetic algorithm runs were carried out using a population size of 150 with each set to carry out 25 million energy evaluations. The results were overwhelmingly localised either near helix E or at the interface between helix E and H, though with quite varied energy values. Lacking any further experimental details it was concluded that the low energy pose from the lowest energy cluster would be the most suitable for further simulation. The selected docked complex is shown in Figure 6.1. This structure was then prepared for classical MD similarly to the protocols outlined in Section 5.2.

6.2.2 Molecular Dynamics Protocols and Outcomes

Several models were considered and can be grouped into wild-type and mutated models. In all cases high affinity bound calcium ions were included and remained bound for the duration of the simulations. The wild-type group consisted of cCTnC·TnI(34-71) modelled alone and with the docked EGCg. The mutated group were the same except the cCTnC residue 159 was mutated from a Glycine to an Aspartic acid. This is referenced as previously by exchanging wild-type with mutated (using the label M_{mut}) in the following work. The focus here was primarily to allow for comparisons of protein-protein interaction which is crucial for the calcium signalling response in the cardiac system: in fact any changes here may affect the muscle response in the full fibre system. The simulations were carried out with AMBER10 using the ff03 force field.[94, 105, 139]

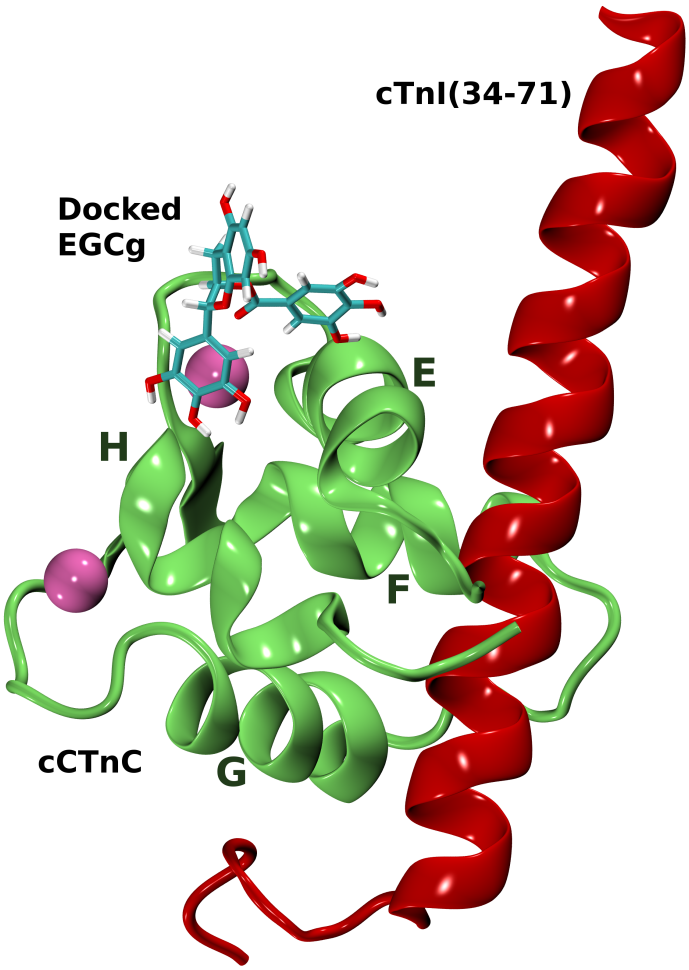

Figure 6.1: cCTnC·TnI(34-71) with the EGCg ligand computationally docked to the E helix and near the E and H helix interface. The cCTnC remains in green while the cTnI fragment is in red.

Preparation steps were carried out in *xleap*, 5 Na^+ counter-ions were placed around the structure in order to neutralise the system using a Coulombic potential on a grid, followed by explicit solvation in a periodically repeated 12 Å buffer truncated octahedral box of TIP3P water molecules. Minimisation and then equilibration of the structures was carried out followed by MD as explained below. PMEMD was again used to carry out the simulations and was run on HECToR, the UK's National supercomputing service as discussed in Section 5.3.

Initially, the solvent and counter-ions were minimised while positional

restraints were placed on the protein and ligand atoms. 20,000 minimisation steps were carried out, the first 15,000 using steepest descent minimisation followed by 5,000 conjugate gradient minimisation steps. Constant volume and periodic boundary conditions were used along with a cutoff of 10 Å for the non-bonded interactions. Second stage minimisation was to relax the entire system and so the positional restraints on the docked protein and ligand atoms were removed. Again 20,000 steps were carried out and divided between steepest descent and conjugate gradient.

MD simulation with harmonic restraints placed on the solute was then carried out to heat the system from 0 K to 300 K using the Langevin thermostat (with a collision frequency of 1.0 ps^{-1}) and at constant volume for 20 ps. The restraints were then removed and the simulation was equilibrated at 300 K for 100 ps. The SHAKE algorithm was also used to constrain the bonds with hydrogen in the protein allowing a time-step of 2 fs.

Production runs of all the models followed, with 50 ns of MD at the constant temperature of 300 K (again using the Langevin thermostat) and constant pressure of 1 atm, using the Berendsen barostat with a time constant of 2 ps. The first 5 ns was considered as the equilibration phase while statistical data were collected from the last 45 ns of simulation. The full simulations were then each repeated twice in order to better assess any unusual effects.

6.2.2.1 Protein Stability for Wild-type and Mutated Cardiac Systems

EGCg remained bound throughout all the simulations however in every case within the first 5 ns the ligand rotated toward the H helix of cCTnC forming hydrogen bond interactions with both the E and H helices where the ligand remained until the end of the simulations. Hence this period was not used to collect data averages. This equilibrated position was seen to be close to both the two major docking poses discussed in Subsection 6.2.1 which is likely to be accessible due to the solvent and, it may be speculated, represents an amalgam of favourable interactions of both the earlier docked poses.

RMSD breakdowns for every model may be seen in Table 6.1. With

Table 6.1: cCTnC·TnI(34-71) models (two replicas) with averaged protein backbone RMSD values along with protein sub-unit breakdowns.

Model	cCTnC	cTnI(34-71)	Total
cCTnC·TnI(34-71)	1.79 ± 0.20	3.06 ± 0.51	2.78 ± 0.42
"	1.84 ± 0.49	2.46 ± 0.59	2.40 ± 0.54
cCTnC·TnI(34-71)·EGCg	1.82 ± 0.20	2.66 ± 0.46	2.58 ± 0.48
"	1.61 ± 0.33	2.57 ± 0.54	2.47 ± 0.52
M_{mut}·TnI(34-71)	2.00 ± 0.23	2.86 ± 0.64	2.79 ± 0.47
"	2.04 ± 0.32	3.24 ± 0.59	3.08 ± 0.64
M_{mut}·TnI(34-71)·EGCg	1.41 ± 0.16	2.81 ± 0.40	2.38 ± 0.33
"	1.72 ± 0.25	2.95 ± 0.62	2.62 ± 0.43

the cTnI(34-71) bound to the hydrophobic cleft, the structure of cCTnC was seen to be highly stable during the simulations as may be expected due to its conformational restriction with the TnI fragment blcoking the hydrophobic cleft. The cTnI(34-71) fragment however was quite mobile on either terminal strand (defined as the last 8 residues on either side of the cTnI fragment). This may be attributed to the unnatural termination of this sub-unit leading to decreased stability in these regions, though the protein remained well bound throughout the simulation time. This instability is exemplified in Figure 6.2 where the cCTnC·TnI(34-71) and cCTnC·TnI(34-71)·EGCg replica 1 models show the backbone fluctuations of the two protein fragments.

Averaging over the eight residues and the eight MD simulations (with and without EGCg and with and without the GLY159ASP mutation) sheds some light on the source of the backbone fluctuations. The C terminal strand of the cTnI fragment was seen to have fluctuations of 2.87 ± 0.58 Å; the largest fluctuations for this terminal were observed in the mutated model M_{mut}·TnI(34–71)·EGCg simulations, without which the average value reduced to 2.70 ± 0.30 Å. The N terminal strand of the cTnI peptide had on average fluctuations of 2.68 ± 0.67 Å; in this case the fluctuations were larger when EGCg was present and for the mutated systems with respect to the wild type.

The interactions of EGCg with the protein complex were primarily through hydrogen bonding, though some $\pi - \pi$ interactions were observed.

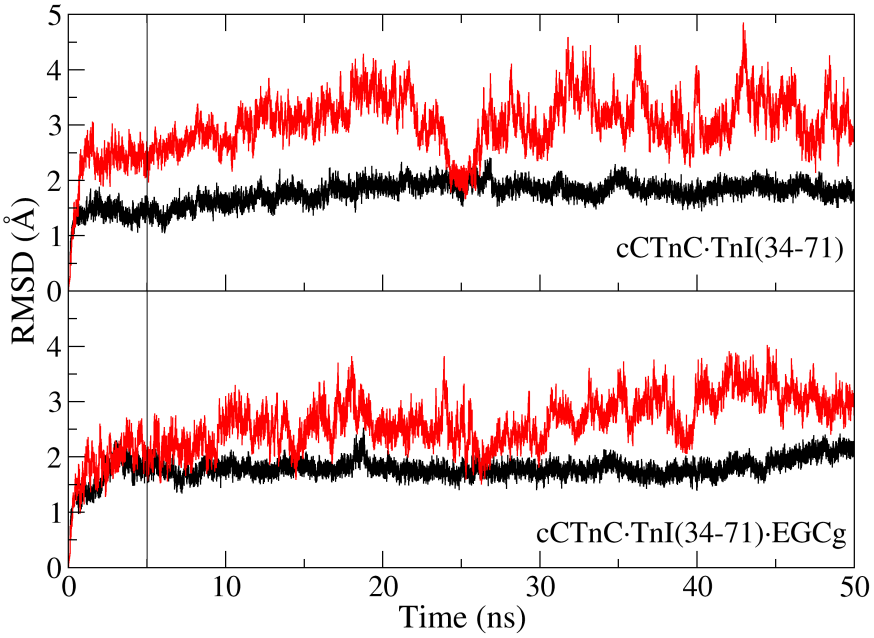

Figure 6.2: Protein backbone RMSDs of the cCTnC·TnI(34-71) complex during simulation broken down by protein sub-unit. In black is the cCTnC while in red is the TnC-(34-71) fragment. On top is replica 1 of the cCTnC·TnI(34-71) simulation while on the bottom is the cCTnC·TnI(34-71)·EGCg replica 1. The drop line shows the point at which the simulations were considered equilibrated and statistics are collected.

EGCg formed on average 19.2 hydrogen bonds of which 4.4 were with cCTnC and 14.8 with the solvent. This can be compared with the earlier catechin work in Table 3.1 where solvation alone accounted for an average 16.9 hydrogen bonds. The increased number of overall hydrogen bonds may be due to the locally favourable environment, in that the EGCg blocks a small hydrophobic patch of the protein and spans two regions where acidic and basic protein residues are likely to increase local density of waters. This combinatorial effect may led to larger clustering of waters near the hydroxyl groups.

Protein-ligand hydrogen bonds were observed with the hydroxyl groups of the B and B′ rings, which considerably restricted the torsion of the

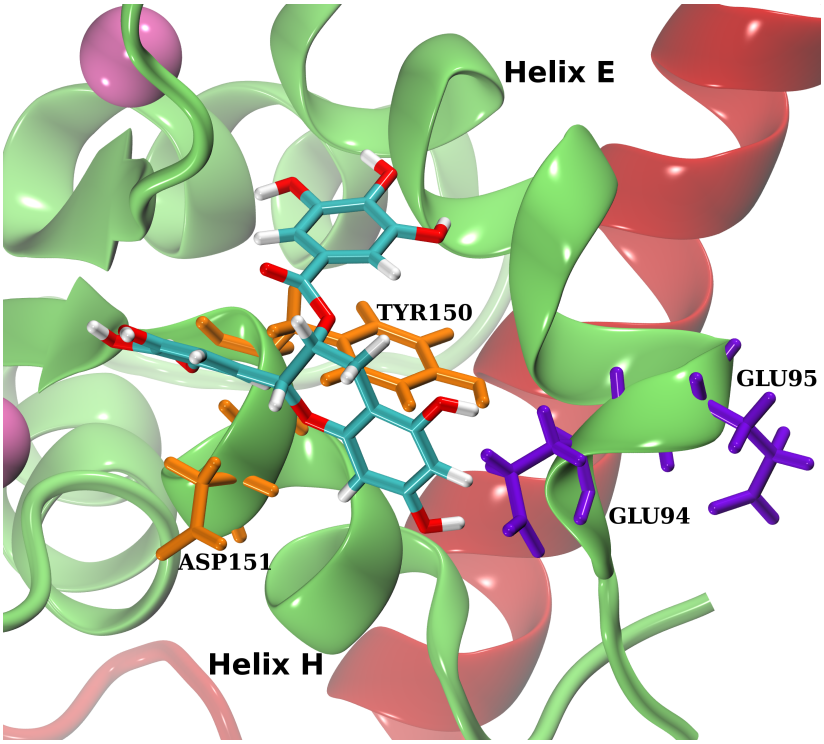

Figure 6.3: MD snapshot after 5 ns of the cCTnC·TnI(34-71)·EGCg replica 1 simulation. Close up view the E - H helix interface displaying several major interactions between EGCg and cCTnC during simulation. In orange are the residues ASP151 (interacting with EGCg ring B here) and TYR150 (π stacking with the B′ ring on helix H). In violet are the GLU94 and GLU95 residues on the E helix with GLU94 currently H-bonding with EGCg ring A.

rings and acted to anchor the EGCg molecule to the E and H helices of cCTnC. Hydrogen bonds between EGCg and cCTnC were formed with the side-chains of the negatively charged residues GLU94 (on average 0.9) and GLU95 (2.3) on helix E as well as ASP151 (1.2) on helix H as shown in Figure 6.3. In the presence of the G159D mutation EGCg formed an average of 16.7 hydrogen bonds, of which 13.3 with water and 3.4 with cCTnC, through a number of residues also including GLU94 (2.05) and ASP151 (0.41), besides SER98 (0.24), ASP105 (0.12), ALA108 (0.11) and TYR150 (0.44). 0.4 $\pi-\pi$ interactions were formed by EGCg with cCTnC·TnI(34–71)·EGCg and 0.5 with M_{mut}·TnI(34–71)·EGCg, in both cases with the TYR150 on helix H which can also be seen in Figure 6.3. These were from both the B and

Table 6.2: Average number of the inter-protein hydrogen bonds and salt bridges between cCTnC and TnI(34-71) (wild type and with the G159D mutation) in the presence and absence of EGCg.

Model	Hydrogen Bonds	Salt Bridges
[cCTnC]·[TnI]	11.2	8.8 (78%)
"	10.5	7.8 (74%)
[M_{mut}]·[TnI]	15.9	11.5 (72%)
"	15.6	11.3 (72%)
[cCTnC·EGCg]·[TnI]	10.6	8.0 (75%)
"	10.6	8.5 (80%)
[M_{mut}·EGCg]·[TnI]	12.5	8.6 (69%)
"	12.5	9.5 (76%)

B′ rings suggesting a weak overall hydrophobic interaction with the ligand.

As shown in Table 6.2, cCTnC and the cTnI(34-71) fragment interacted via several strong hydrogen bonds the majority of which were salt bridges between amino acid pairs of opposite charge. The observed dominance of salt bridges is consistent with the experimental suggestion that cTnC-cTnI binding is electrostatically driven[186]. The total average number of hydrogen bonds was increased with the mutation of GLY159 with the negativly charged ASP, as we would intuitively expect. However the increased number of hydrogen bonds in the mutated systems, which led to stronger binding, seems to be in contrast with NMR experiments which detected a slight weakening.[31] Overall, the presence of EGCg reduced or maintained the average number of hydrogen bonds, which may be related to a weakening of the interaction between cCTnC and cTnI(34-71).

6.2.2.2 Calcium Coordination in the presence of cTnI

The coordination number of the two calcium ions with oxygen was calculated in the same manner as described in Sub-Section 5.2.1 and the results are shown in Table 6.3.

In the same manner as for the cCTnC simulations in Chapter 5 both the calcium ions coordinate between 8 and 9 nearby oxygens. The site 2 calcium shows extremely low variation across simulations and replicas and appears to readily coordinate with 9 oxygens. It appears that the site 1

Table 6.3: The coordination numbers of calciums ions averaged throughout the cCTnC·TnI ff03 MD simulations.

Model	Calcium 1	Calcium 2
cCTnC·TnI(34-71)	8.6	8.9
"	8.2	8.9
M_{mut}·TnI(34-71)	9.1	8.9
"	8.8	9.0
cCTnC·TnI(34-71)·EGCg	9.0	9.0
"	8.9	9.0
M_{mut}·TnI(34-71)·EGCg	8.9	9.0
"	9.1	8.9

calcium shows higher overall coordination than previously except in the case of the cCTnC·TnI(34-71) replicas. This overall slight increase in the site 1 value could be explained by the reduction in protein backbone fluctuations leading to a slightly tighter binding overall, however the difference is quite small compared to cCTnC alone and these results do not seem to change the picture of calcium coordination in the system.

6.2.2.3 Binding Free Energy of the cCTnC·TnI(34-71) Complex

The binding free energy of the protein-protein interaction and protein-ligand interaction was evaluated using the previously discussed MM/PBSA method with AMBER. Both the enthalpic and entropic contributions to binding were calculated and are shown in Table 6.4. As previously discussed in Chapter 5, a dielectric constant of 2.0 was used to account for the high content of positively and negatively charged amino acid residues in the protein (notably cTnI(34-71)) leading to a large percentage of interactions between the protein sub-units consisted of salt-bridges as shown in Table 6.2.

As before, the entropic contributions remained largely similar between systems while the enthalpic contribution showed the largest variations. The $\Delta\Delta$G values were calculated with respect to the wild-type system allowing comparative discrimination between models. Little overall difference is seen between the binding of EGCg to the wild-type or mutated protein suggesting that the catechin propensity to bind would not be affected by

Table 6.4: Enthalpic and entropic contributions and binding free energies between the complexes in square brackets evaluated within the MM/PBSA and normal mode analysis scheme. Units are in kcal/mol.

Model	ΔH	$-\mathrm{T}\Delta S$	ΔG	$\Delta\Delta G$
[cCTnC·TnI]·[EGCg]	-9.4 ± 2.5	15.8 ± 2.5	6.5 ± 3.5	0.0
"	-11.6 ± 3.6	17.9 ± 2.0	6.3 ± 4.1	-0.2
[M_{mut}·TnI]·[EGCg]	-14.0 ± 3.9	18.1 ± 3.9	4.1 ± 5.6	-2.4
"	-9.1 ± 4.4	16.2 ± 3.6	7.0 ± 5.7	0.5
[cCTnC]·[TnI]	-129.7 ± 8.4	55.5 ± 5.9	-74.2 ± 10.3	0.0
"	-130.1 ± 8.1	56.2 ± 5.5	-73.9 ± 9.8	0.3
[M_{mut}]·[TnI]	-139.3 ± 10.1	55.8 ± 6.3	-83.5 ± 11.9	-9.3
"	-137.0 ± 11.7	53.5 ± 5.6	-83.4 ± 13.1	-9.2
[cCTnC·EGCg]·[TnI]	-121.0 ± 10.7	55.1 ± 5.1	-65.8 ± 11.9	8.4
"	-123.1 ± 9.9	52.8 ± 5.8	-70.3 ± 11.4	3.9
[M_{mut}·EGCg]·[TnI]	-126.4 ± 12.5	56.7 ± 5.7	-69.7 ± 13.7	4.5
"	-121.6 ± 6.9	52.6 ± 5.6	-70.0 ± 8.9	4.2

this mutation. Of more interest is the effect of mutation and EGCg binding on the protein-protein interaction. There is a clear trend that mutation increases the strength of the binding between protein sub-units while the addition of EGCg weakens it. Crucially, the catechin weakening effect is still seen in the mutated system to a smaller degree lending credence to the idea of EGCg as a 'moderator' of the interaction. As this interaction passes on the calcium binding signal in the full protein system this may prove useful as the moderating effect could return more normal function to the diseased state.

6.3 Revisiting Experimental Data: Two Potential Binding Sites

At this point, a re-examination of the experimental literature is instructive. The work by Sykes et al.[6] was able to produce several bound structures for the cCTnC·EGCg system which were then verified and expanded upon by our work in Section 5.2. At the time the experimental group had also attempted to investigate the cCTnC·TnI(34-71) protein complex but were

unable to fully make sense of the chemical shifts in the NMR data. They did however suggest that a binding site might exist at or near the F helix of cCTnC in this case. The work in Section 6.2 was able to shed some light onto this system and explained some of the chemical shifts seen in the region near helix E and H. Close inspection of this area as well as that posited by the experimentalists suggested that these two regions were in fact very similar and a small number of chemical shifts are observed in both regions. Both sites are at a helical interface, E & H and F & G, and have small hydrophobic patches exposed to the exterior of the protein with charged residues on either side. It was realised that the reason for the unusual NMR data might in fact be due to there being two binding sites for EGCg on either side of the cCTnC sub-unit.

Using the docking and MD data gathered from Sections 5.2 and 5.3 as a template, as well as the chemical shift data previously obtained[6], the experimentalists were able to produce a double bound structure using *xplor-nih*[187, 188] (a structure determination program which uses chemical shift data to produce bound structures) to dock two EGCg molecules to cCTnC·TnI(34-71). The backbone protein atoms were kept rigid while side-chains were allowed to rotate and the two ligands were docked using a simulated annealing protocol. In the case of the second site (the one not found with our computational docking earlier) there was some ambiguity in the assignment of NOE shifts and so the possibility for both restraints to troponin C residues LEU114 and ILE119 were included. It is hoped that through MD it will be possible to determine the correct preference, if any. The lowest energy structure obtained from this method was extracted is shown in Figure 6.4. This was then prepared for MD to test the validity of this model and whether the molecules would remain bound concurrently or whether there were simply two sites which could be occupied.

6.3.1 Molecular Dynamics Protocols

6.3.1.1 New Force Field Testing

Molecular modelling is a field which is continually being refined and improved. As such, new force fields are often added to revisions of computing

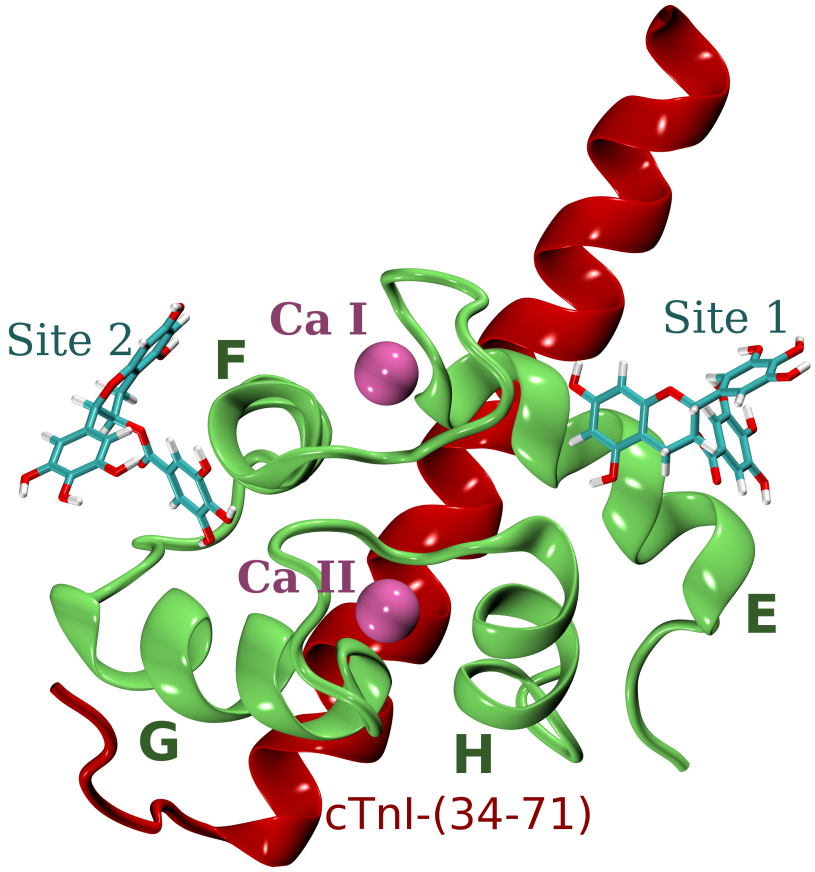

Figure 6.4: cCTnC·TnI(34-71) with lowest energy NMR double docked EGCg displaying labelled sites, calcium ions and helix nomenclature.

codes and existing force fields are updated with time to include new empirical parameters derived either from experiment or high level quantum simulations. Before further production simulations were carried out on the double-bound troponin system, several newer force fields were tested for their suitability to represent the system in question. As had been seen in Section 5.3, the use of a peptide fragment, cTnI(34-71), had led to small instabilities at the terminal strands. This had a knock on effect on several aspects of the simulations such as the fluctuations of the binding free energy. Unfortunately this problem could not be easily rectified by simply including a larger portion of the cTnI protein as this has not been completely experimentally resolved. This is not to suggest that the system is impossible to study however, but it behooves us to investigate whether improvements

Table 6.5: Protein structure stability based on backbone RMSD of the protein and compared with defined secondary structures.

Force field	Total (Å)	Secondary Structure (Å)
ff03	2.78 ± 0.42	1.59 ± 0.30
ff03r1	2.59 ± 0.36	1.85 ± 0.31
ff99SB-ildn	2.05 ± 0.37	1.79 ± 0.27
ff12SB	2.24 ± 0.34	1.52 ± 0.24

to protein force fields can partially enhance current data. With this in mind several shorter classical simulations were undertaken using a range of force fields, from the well known and tested to the new and 'improved' to determine a reasonable choice for further calculation. The selected force fields were the revised ff03r1, the ff99SB with ildn correction and the newer ff12SB.[189] A brief discussion of their parameter derivation and major differences is undertaken in Section 2.3, Subsection 2.3.1. Simulations of the cCTnC·TnI(34-71) complex were carried out using the equilibration steps outlined in Section 6.2, Subsection 6.2.2. The production run was set to be 50 ns in order to collect comparable statistics to the simulations run with the ff03 force field and the last 45 ns of this was used for statistical averages as before. The backbone RMSD averages for each force field used during the production run can be seen in Table 6.5. The overall differences between the studied force fields were small and in the case of the secondary structures tended to fall within the standard deviations, though the ff99SB and ff12SB did uniformly show increased secondary structure stability over the ff03 based simulations. The average fluctuations were also marginally decreased in the ff99SB and ff12SB simulations suggesting better helical and beta strand behaviour in these cases. This appears to show that the modifications to the side chain angles have led to a reduced overall structural fluctuation. Due to the combination of improved total structural stabilisation and reduced overall fluctuations the 12SB force field was chosen for later MD work on the complex.

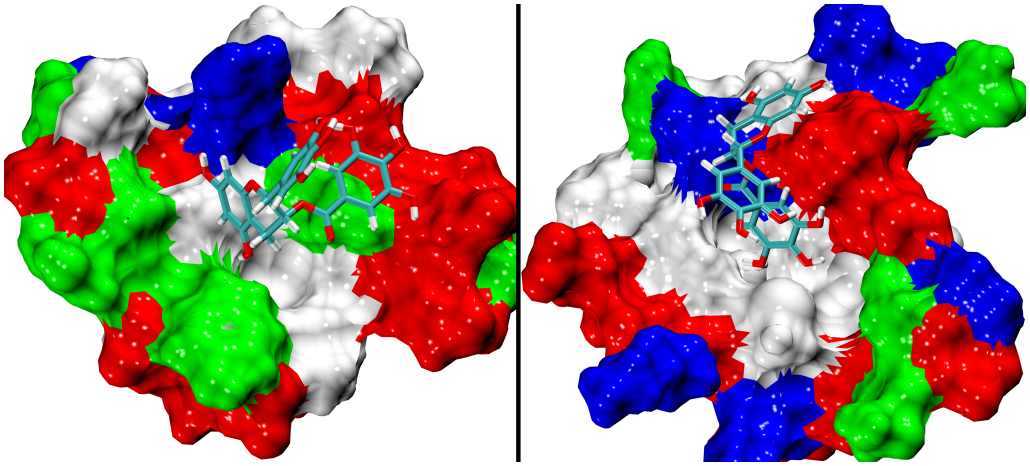

Figure 6.5: Close up view of the Site 1 and Site 2 docked locations generated from experiment including the protein molecular surface coloured by residue type: Red - acidic, blue - basic, green - polar un-charged, white - hydrophobic. The two sites share a number of similarities including the presence of hydrophobic residues partially buried in the interface between helices of charged or polar residues.

6.3.2 Model Preparation and Simulation Protocols

The cCTnC·TnI(34-71) structure with two bound EGCg molecules provided the basis for the simulations undertaken. Figure 6.5 shows close views of the binding sites of this structure with a view of the local protein environment. Four protein complex systems were considered, representing the possibilities of the docked double bound complex: the cCTnC·TnI(34-71) complex to represent the original system; cCTnC·TnI(34-71)·EGCg$_{S1}$; cCTnC·TnI(34-71)·EGCg$_{S2}$ and cCTnC·TnI(34-71)·EGCg$_{DB}$. Here and from this point on the EGCg$_{S1}$ refers to the binding site at the helical E & H interface (Site 1), EGCg$_{S2}$ represents the binding site at the helical F & G interface (Site 2) and EGCg$_{DB}$ represents the double bound system. The protein complexes were prepared with *xleap* in a manner similar to that discussed previously with 12 Å of explicit TIP3P waters and counterions placed to fully neutralise the entire system.

MD preparation of the models also followed previously established patterns with minimisation using protein-ligand positional restraints followed by full minimisation of the system. Equilibration was carried out with NVT

for 20ps with protein-ligand positional restraints while heating the solvent to 300K was undertaken followed by a further 100ps of NPT equilibration without restraints using the Langevin thermostat (with 1 ps^{-1}) and Berendsen barostat as has been established in earlier Chapters. Full production runs for 50ns of NPT simulation were then carried out with statistics collected from the final 45 ns. These simulations were each repeated 3 times in order to generate statistical averages for the systems.

6.3.3 Single-Bound EGCg

The two single bound complexes are considered here separately due to their differing locations and subsequent interactions. The $EGCg_{S1}$ system is representative of the earlier, computationally docked system investigated in Section 6.2 and as such some comparisons may be drawn between the two though force field effects must be taken into account. The $EGCg_{S2}$ system has not been previously studied and so represents an alternative possibility which has not been explored.

6.3.3.1 Site 1

As expected from earlier simulations, the three replica $EGCg_{S1}$ models all show EGCg to remain bound during the 50 ns of simulation. In addition the protein backbone and secondary structure remain stable throughout.

As the $EGCg_{S1}$ system is effectively very similar to the earlier computationally docked model in Section 6.2 it is not surprising to note that a number of the same interactions with EGCg are seen, notably both H-bonding and $\pi - \pi$ bonding between EGCg and TYR150. The differences here are likely due to the effects of the new force field and slightly different initial positions. The major ligand - protein hydrogen bonds which were formed during the replica simulations can be seen in Table 6.6. A bond is included in this table if it was found to exist for more than 10% of the total production run. The bond listed does not differentiate between cases where the EGCg hydroxyl group acts as donor or acceptor, only that the bond existed, for brevity.

Listed in the table are the protein and solvent H-bond totals as well

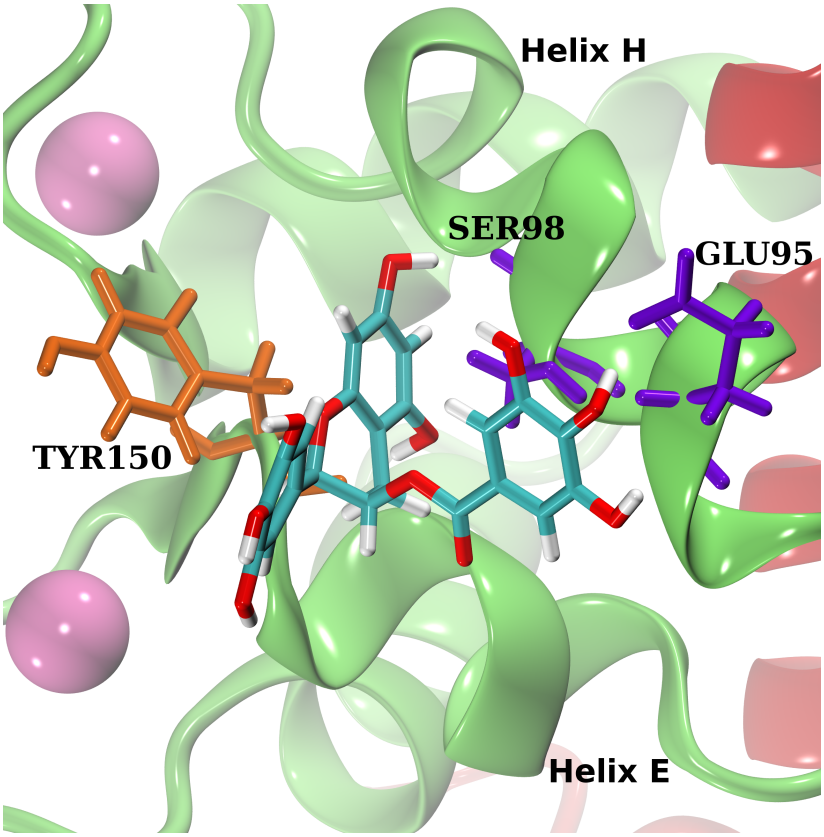

Figure 6.6: Replica 1 of the Site 1 simulation at the end of the equilibration phase showing interacting residues. In orange is the TYR150 on the H helix while in violet is the GLU95 and SER98 on the E helix. This is illustrative of the EGCg A ring sliding between the E/H interface and interacting with both sides.

Table 6.6: $EGCg_{S1}$ Major EGCg hydrogen bonds averaged throughout replica simulations.

Replica	Protein Residue (Helix)	Protein H-bonds	Water H-bonds	Total H-Bonds
1	GLU95, SER98 (E) TYR150 (H)	2.7	12.5	15.2
2	GLU94, SER98, ARG102 (E) TYR150 (H) ALA108 (E/F loop)	3.1	10.9	14.0
3	GLU94, LEU97 (E) ASP109 (E/F loop)	2.0	12.9	14.9

Table 6.7: Site 1 $\pi - \pi$ Interactions during production runs between EGCg and local PHE and TYR residues.

Replica	Protein Residue	Total
1	TYR150	66%
2	TYR150	15%
3	TYR150	81%
	PHE101	4%

as secondary structure helix location of the protein residues in brackets (in some cases the residue did not belong to a specific helix and instead the loop region location between helices is instead listed). Figure 6.6 displays the interactions of the Replica 1 simulation at the end of the 5 ns equilibration phase. The major H-bonding residues are shown as well as the TYR150 currently interacting via π stacking with ring A. This is very similar to the picture presented in Figure 6.3 with the exception of ring orientation. Table 6.7 details the total $\pi-\pi$ interactions which occurred during the simulations between EGCg and cCTnC. The criteria for $\pi-\pi$ bonding is detailed clearly in Chapter 5, Section 5.3.

Some $\pi-\pi$ bonding exists in all the simulations, almost exclusively with TYR150 but a small amount with PHE101 as well. The EGCg B$'$ is not seen to interact in this manner in any case. In the first and third replica a $\pi-\pi$ bond is seen to exist for a significant portion of the production run and this strongly anchors the ligand to the H helix in both cases. In the second replica this bonding is reduced and appears to be compensated for by a stronger hydrogen bond to the TYR150 instead indicating stability in both the parallel and perpendicular ring orientations. This allows the EGCg molecule to remain anchored to the H helix while the B$'$ ring maintains hydrogen bonds with several residues on the E helix. Tables 6.6 and 6.7 also indicate a distinct effect of π bonding in that, due to the conformational restrictions when in such a bond, the EGCg hydroxyl groups are less likely to form H-bonds with protein residues. These other hydrogen bonds therefore appear weaker overall than those with TYR150, which is likely due to the high availaibility of suitable protein residues rather than indicating a more loosely bound system. In replica 3 EGCg appears to briefly form between

$\pi - \pi$ interactions to the nearby TYR150 and PHE101 as well as forming short-lived hydrogen bonds with E/F loop residues, the most prevalent of which appeared to be to ALA108. Overall this provides an interesting insight into the strength of the $\pi - \pi$ bonds occurring in the system.

Using MM/PBSA methods, estimates for the enthalpic and entropic contributions to the free energy of binding for the system were carried out. The enthalpic portion of the free energy was calculated with the Poisson-Boltzmann methodology on 4,500 frames of the replica simulations. The entropic contributions were carried out with normal mode calculations on 20 frames taken from the replica simulations at evenly spaced intervals with nmode in AMBER. The external dielectric was set to 80 to mimic water solvation and the internal dielectric was set to 2 as we had previously shown this to be a better fit to model binding behaviour in this system. As has been discussed before, the results of these calculations should be used to determine relative binding strength of similar molecules rather than an absolute value. It should also be noted that as a newer force field is being used these results are not directly comparable to the earlier results in Table 6.4 though we are still able to draw trends from the data. The results of the calculations are displayed in Table 6.8.

From the earlier investigation on local interactions throughout simulation we can say it is likely that the large entropic value for replica 2 is due to a less favoured ring orientation deep within the protein (the lack of strong $\pi - \pi$ bonding and the stronger H-bond to TYR150 suggest a preference for parallel orientation between aromatic rings). In replica 3 the lower number of H-bonds is likely correlated with an increased in π stacking with the TYR150 and compared to replica 1 appears to be the cause of the more favoured binding.

Overall we see a favourable ΔG value with a consistent binding energy across all replicas, though the slightly different interactions of replica 3 suggest the site is not rigidly bound.

6.3.3.2 Site 2

The new site for EGCg binding has shown to be stable throughout the simulation time similarly to the EGCg$_{S1}$ system. Unlike in the EGCg$_{S1}$

Table 6.8: $EGCg_{S1}$ MM/PBSA Calculation Results Showing Enthalpic and Entropic Contributions for the protein-ligand binding using an internal dielectric of 2.0.

Replica	ΔH	$-T\Delta S$	ΔG
1	-19.3 ± 3.1	15.8 ± 4.7	-3.5 ± 6.1
2	-22.4 ± 3.5	20.1 ± 4.2	-2.3 ± 5.5
3	-20.4 ± 4.3	16.8 ± 5.0	-3.6 ± 6.6
Average	-20.7 ± 3.6	17.5 ± 4.8	-3.7 ± 6.1

situation the location of $EGCg_{S2}$ prevents any possible $\pi - \pi$ interactions with EGCg as the TYR and PHE residues of cCTnC are far from the new potential binding site.

Thus a major stabiliser of the bound system is hydrogen bonding, though reduction of solvent exposure to the nearby MET137 and hydrophobic LEU114 residues may also be a factor to the favourability of binding. While this is the preferred site from the experimental NMR analysis[6] due to slightly stronger chemical shifts, it was still a tentative suggestion and the lack of clear NOESY signals initially indicates that this could be a transient binding. Comparison of the replica simulations shows that the EGCg does not dissociate during simulation time and seems to suggest that while hydrogen bonding occurs frequently, these are usually seen to be shorter-lived bonds which the EGCg hydroxyl groups frequently switch between. A breakdown of the major contributors is shown in Table 6.9.

In replica 1 the majority of these interactions appear to exist between the EGCg A and B′ rings to residues on helix G on one side while a strong hydrogen bond between ring B and helix F anchors the ligand to the binding site on the other. In replica 2 the H-bonds are primarily to the G helix with a slightly weaker bond to helix F from the B′ ring. In replica 3 the EGCg benzopyran A/C rings appear to interact more with the F helix while the B and B′ rings maintain contacts to the G helix, though they are all in general quite weak interactions compared to the other replicas. In each replica it appears that the total hydrogen bonds of the A and B′ rings are nearly identical. This seems likely to increase the overall stability of the system due to the ability to compensate for local system variations, though it may

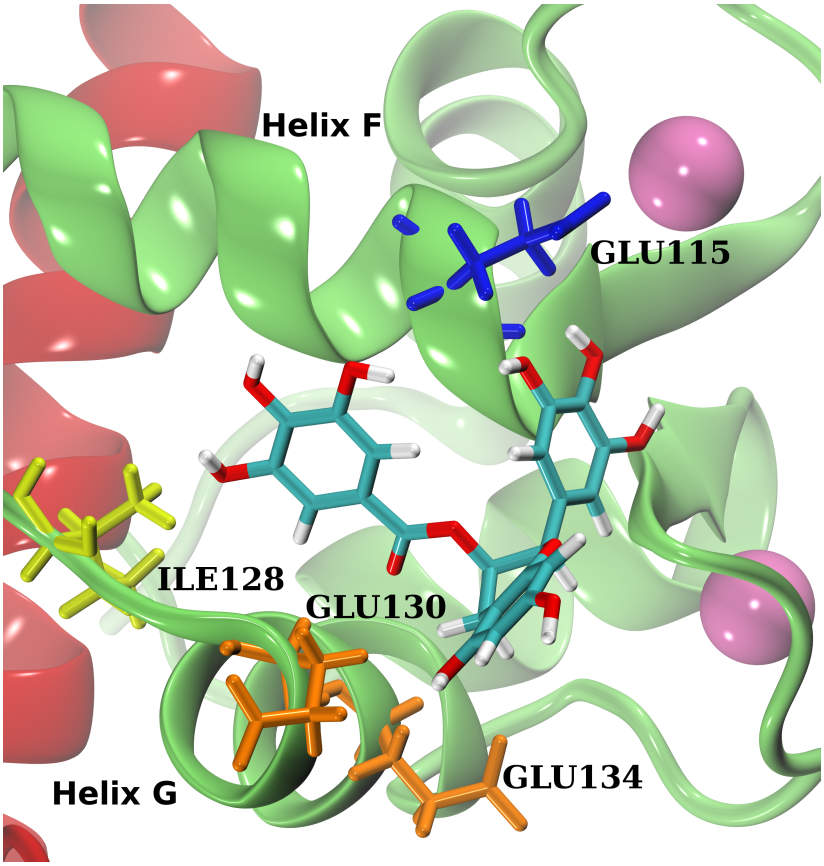

Figure 6.7: Close up view of the protein-ligand interactions of Replica 1 of the Site 2 simulations after equilibration. In orange are the GLU130 and GLU134 residues on the G helix, in blue is the GLU115 on the F helix and in yellow is the ILE128 residue on the F/G loop.

indicate that the binding is short lived or transitory. As seen in the earlier Chapter 3 the B′ ring has larger energy barriers to torsion than the B ring and thus it is expected that this imposes some difficulty in achieving a close bound structure with EGCg as the ligand interactions are unlikely to match up precisely with the individual ring energy minima. However in this case the A/B′ rings stabilise one side while the B ring flexibly interacts on the other, leading to quite variable binding strength.

MM/PBSA was performed for estimates of the free energy of binding using a method identical to the EGCg$_{S1}$ method. The results can be seen in Table 6.10. Unlike the EGCg$_{S1}$ case, here there appears to be quite large

Table 6.9: $EGCg_{S2}$ Major EGCg hydrogen bonds averaged throughout replica simulations.

Replica	Protein Residue (Helix)	Protein H-bonds	Water H-bonds	Total H-Bonds
1	GLU130, GLU134 (G) GLU115 (F) ILE128 (F/G loop)	2.0	13.1	15.1
2	ILE133, GLU130, GLU134 (G) LEU114 (F) GLY146 (G/H loop)	2.5	10.9	13.4
3	MET137, LYS138 (G) ASP113, GLU115 (F) ASN144, GLY146 (G/H loop)	2.5	11.8	14.8

Table 6.10: $EGCg_{S2}$ MM/PBSA Calculation Results Showing Enthalpic and Entropic Contributions for the protein-ligand binding using an internal dielectric of 2.0.

Replica	ΔH	$-T\Delta$S	ΔG
1	-15.0 $\pm$ 3.7	15.7 $\pm$ 4.7	0.7 $\pm$ 6.0
2	-26.9 $\pm$ 2.9	17.2 $\pm$ 5.4	-9.8 $\pm$ 6.1
3	-17.5 $\pm$ 5.5	21.7 $\pm$ 5.9	4.2 $\pm$ 8.0
Average	-19.8 $\pm$ 4.1	18.2 $\pm$ 5.3	-1.6 $\pm$ 6.7

variations in both contributions to the free energy of binding. Enthalpic contributions are much lower except in the replica 2 case. Entropically, replicas 1 and 2 give similar values to the $EGCg_{S1}$, replicas 1 and 3 while replica 2 here is much larger. This is due to the replica 3 B′ ring orientation facing deeply into the protein helix cleft (where it interacts strongly with ILE133, not seen in the other replicas) causing a large increase in the vibrational contribution to entropy for the complex.

While the average binding remains favourable overall, there is a large variation in replica values. This may be due to the fact that EGCg in $EGCg_{S2}$ is more likely to interact with charged residues, with fewer nearby hydrophoic possibilites. It could also be due to the uncertainty in the initial experimental structure determination leading to a poorer binding. Replica

Table 6.11: $EGCg_{DBs1}$ Major EGCg hydrogen bonds averaged throughout replica simulations.

Replica	Protein Residue (Helix)	Protein H-bonds	Water H-bonds	Total H-Bonds
1	GLU95, SER98 (E) GLU155 (H)	3.7	12.4	16.1
2	GLU95, ASP99, ARG102 (E) ASP105 (E/F loop)	3.7	12.1	15.8
3	GLU95, SER98, ASP99 (E) TYR150 (H)	1.6	13.4	15.0

2 is the only replica which shows direct hydrogen bonding to the residues LEU114 or ILE119 used in the original *xplor-nih* structure determination which may partially explain its increased enthalpy. The difference between $EGCg_{S1}$ and $EGCg_{S2}$ binding favourability here appears quite small but noticable, around 2 kcal/mol, almost entirely due to the enthalpic term which may indicate a binding site preference.

6.3.4 Double Bound EGCg

The final set of replica simulations generated were for the doubly bound case where an EGCg molecule is attached to both sites from the outset. Here the $EGCg_{DBs1}$ and $EGCg_{DBs2}$ nomenclature is used to refer to each in the combined system. For all three replica simulations both the site 1 and site 2 EGCg molecules remained bound during the full production runs. Tables 6.11 and 6.14 give breakdowns of the major protein - EGCg hydrogen bond interactions for the $EGCg_{DBs1}$ and $EGCg_{DBs2}$ locations respectively. Table 6.12 shows the $\pi - \pi$ interactions with the $EGCg_{DBs1}$ in these simulations. As before, $EGCg_{DBs2}$ did not undergo any $\pi - \pi$ bonding due to the binding site location. It is useful to compare the differences for each site in the double bound system with those of the singly bound simulations discussed previously. Due to their location on opposite sides of the cCTnC protein we expect no direct interaction, and indeed we find this is the case, however it is possible that the influence on the protein on one site could cause a protein mediated change in the other leading to a further possible cooperative effect.

Table 6.12: $EGCg_{DBs1}$ $\pi - \pi$ interactions during production runs between EGCg and PHE and TYR residues of cCTnC.

Replica	Protein Residue	Total
1	-	-
2	TYR150	6%
3	TYR150	42%
	PHE101	4%

Inspection of Tables 6.11 and 6.12 immediately reveals several differences from the singly bound system. The overall number of H-bonds tends to be higher (in fact the average number of protein - ligand H-bonds averaged for the single bound case was 2.6 while for the double bound is 3.0), while the prevalence for $\pi - \pi$ bonding seems to have correspondingly decreased. The decrease in $\pi - \pi$ bonding appears particularly stark with replica 1 showing no bonding of this type at all while replica 2 shows only very slight bonding, and with the B′ ring, not seen to interact in this manner at all in the single bound case.

It is probable that these effects are linked, an increase in strong H-bonding would prevent sufficient overlap of aromatic rings which is likely to have occurred here. Very little H-bonding in Replica 1 occured outside of the major bonds listed, around 0.2 H-bonds in total. These few, strong H-bonds appear to have existed for the entire simulation time and this strong anchoring to helix H and E leaves little space for the rings to orient close to the protein surface in a way which allows aromatic stacking. Replica 2 shows a similarly high level of H-bonding, though with weaker H-bonds to GLU95 and ARG102, with only 0.15 bonds accounting for 'minor' interactions. In this case anchoring of the ligand appears to come from the very strong H-bonds to ASP105, a residue buried relatively deep into the E/H helix cleft. This formation then causes EGCg to favour an alignment of the B′ ring to run parallel down the cleft between the helices, leaving a near freely rotating B ring exposed to solvent similar to Replica 1. Replica 3 in the double-bound system appears quite similar to the single $EGCg_{S1}$ replica 3 model, with fewer H-bonds than the others but a considerable amount of $\pi - \pi$ bonding to TYR150. Here the A ring has aligned along the helix

Table 6.13: $EGCg_{DBs1}$ MM/PBSA calculation results showing enthalpic and entropic contributions for the protein-ligand binding using an internal dielectric of 2.0.

Replica	ΔH	$-T\Delta S$	ΔG
1	-17.9 ± 2.5	17.9 ± 4.6	0.0 ± 5.2
2	-17.5 ± 3.3	19.3 ± 3.5	1.8 ± 4.8
3	-18.0 ± 3.6	18.5 ± 4.7	0.6 ± 5.9
Average	-17.8 ± 3.1	18.6 ± 5.0	0.8 ± 5.3

interface anchored to the H helix with B ring H-bond interactions connecting the E helix. Relatively few interactions take place with the B′ ring as this is positioned at the edge of the interface and directs away from the protein for much of the production run.

Estimates for the free energy of binding for the system were carried out using the MM/PBSA methods as described earlier using an internal dielectric of 2.0. The results can be seen in Table 6.13. As explained for the earlier simulations this was carried out in order to more accurately describe the local protein environment which contains many acidic and basic residues in close proximity to both EGCg sites (for example in Table 6.11 the largest number of H-bonds in each replica is formed to a negatively charged GLU or ASP residue). The overall effect of this on the enthalpic contribution appears to show the replicas providing a very consistent value for the favourable binding term. The averaged final result of 0.8 kcal/mol suggests a very slightly disfavoured binding. Interestingly, it appears that the average binding is reduced compared to the $EGCg_{S1}$ simulations by about 1.5 kcal/mol suggesting that the site 1 location is less favourable a binding site when site 2 is already occupied though these differences are still well within the standard deviations of the calculations.

$EGCg_{DBs2}$ H-bonds are shown in Table 6.14 and depict a seemingly less well-defined interaction space than the $EGCg_{DBs1}$ system.

In replicas 2 and 3 several interactions occur which include O11, the double-bonded carbon close to the B′ ring. This is slightly unusual and does not appear in other replicas but is likely to have occurred due to close H-bonding to the O3 hydroxyl which is nearby on the B′ ring itself. In

Table 6.14: $EGCg_{DBs2}$ Major EGCg hydrogen bonds averaged throughout replica simulations.

Replica	Protein Residue (Helix)	Protein H-bonds	Water H-bonds	Total H-Bonds
1	LYS118 (F) GLU134 (G) ASP145, GLY146 (G/H loop)	3.7	10.9	14.6
2	GLU134, LYS138, ASP141 (G)	3.1	10.8	13.9
3	LYS118, GLN122 (F) GLU130 (G) GLU126, ILE128 (F/G loop)	2.8	11.4	14.2

replica 2 this interaction along with the strong H-bonds from the B ring are the only way the ligand anchors to the protein and reveal these two rings align closely to the protein surface while the A ring faces away. Due to the position of EGCg in site 2 and the specific H-bonds formed, the ligand is able to remain in the F/G helix interface without any major F helix interactions. As previously stated, the site 2 location only forms hydrogen bonds with no $\pi-\pi$ interactions taking place. The prevalence for H-bonding in the double-bound case appears to have increased compared to the $EGCg_{S2}$ system forming on average around 3.2 H-bonds versus about 2.3 for the single site. It is possible that this is due to protein-wide conformational shifts due to binding at both sides of the protein. This may additionally explain the fewer $\pi-\pi$ interactions seen in the $EGCg_{S1}$ case. Analysis of the $EGCg_{S2}$ H-bonds against the $EGCg_{DBs2}$ H-bonds suggests that in the single case there is a preference for binding slightly closer to the G/H loop end of the helical interface and thus we see more H-bonds form with these residues. Meanwhile in the double bound case the site 2 EGCg does not seem to have a preference and interactions are seen on either end.

MM/PBSA calculations were carried out as before for both enthalpic and entropic contributions (using normal mode analysis) as has been discussed for the $EGCg_{DBs1}$ and elsewhere. Results are shown in Table 6.15. For all three replicas the entropic terms remained very similar suggesting little difference in orientation of the EGCg around and within the protein. In

Table 6.15: $EGCg_{DBs2}$ MM/PBSA calculation results showing enthalpic and entropic contributions for the protein-ligand binding using an internal dielectric of 2.0.

Replica	ΔH	$-T\Delta$S	ΔG
1	-21.5 ± 2.7	19.2 ± 6.4	-2.3 ± 6.9
2	-20.5 ± 2.8	18.1 ± 5.2	-2.4 ± 5.9
3	-18.5 ± 4.0	18.6 ± 4.3	0.1 ± 5.9
Average	-20.2 ± 3.2	18.6 ± 5.3	-1.5 ± 6.2

Table 6.15 we can see that consistent results are achieved for the enthalpic term, with similar standard deviations across replicas. In this simulation set replica 3 is very similar to the $EGCg_{S2}$ replica 1 simulation and it is possible that the presence of the hydrophobic residue ILE128 leads pbsa to underestimate overall binding of the system.

It is also worth noting that the $EGCg_{S2}$ and $EGCg_{DBs2}$ give nearly identical averages for total ΔG with a dielectric of 2.0. This indicates that Site 2 binding is not affected by EGCg concurrently binding to Site 1 and it may be the preferred site for EGCg binding when there is an excess of EGCg in the cardiac sarcomere.

6.3.5 Calcium Mobility and Protein Stability

This Subsection contains accumulated data and statistics for the earlier simulations which are more easily compared side by side rather than in their respective separate sections.

Figure 6.8 and Figure 6.9 shows relevant RMSD protein backbone data as well as calcium ion fluctuations from the simulations undertaken. It is worth noting that when averaged over all the replicas, the undocked cCTnC·TnI(34-71) backbone fluctuations are very similar to those seen in Section 6.2 suggesting that the choice of a more up to date force field has not markedly changed the protein behaviour, which validates our previous results. Several familiar trends can be picked out from this data, including a slight overall decrease in structural fluctuation seen when there is bound EGCg in the system. This effect is more visible for $EGCg_{S1}$ and $EGCg_{DB}$ and difficult to discern in $EGCg_{S2}$ which may indicate that site 2 alone has

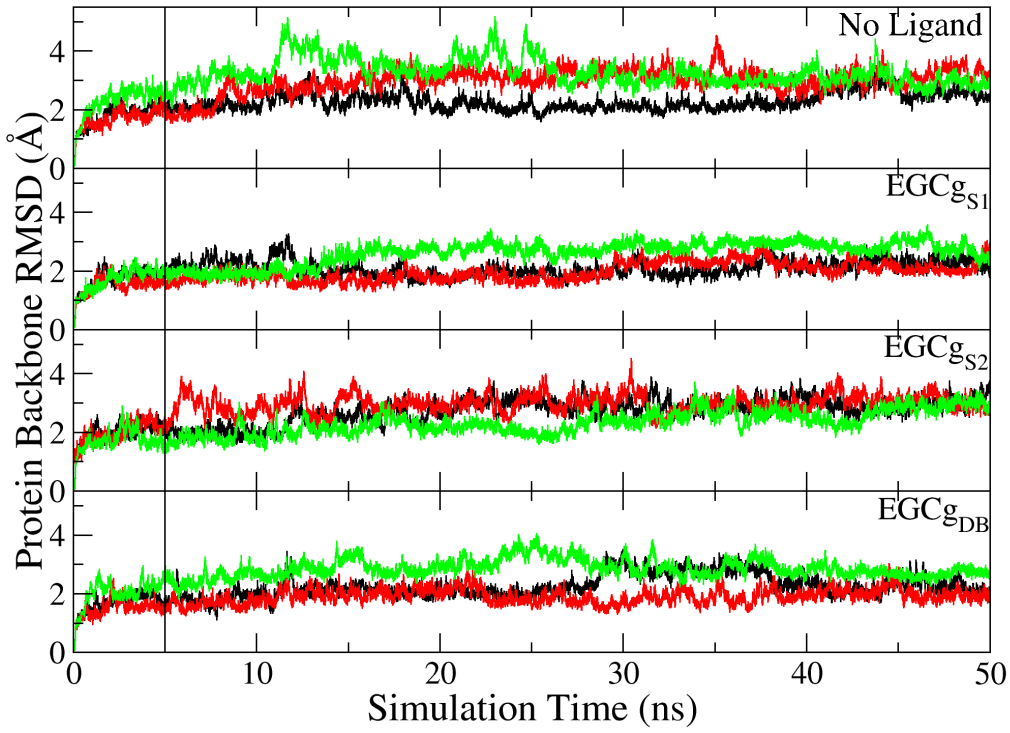

Figure 6.8: Protein backbone RMSDs versus simulation time for each model simulated. In each case the black line is replica 1, red is replica 2 and green is replica 3. The drop line at 5 ns indicates where the production runs began.

little to no effect in enhancing structural stability of the protein. When comparing the calcium ion scatter plots in Figure 6.9 there appears to be some variation between the model replicas, particularly for the unbound complex and the two single site models. It is therefore difficult to discern an overall trend between the models, however, in general the ligated models show slightly reduced calcium ion movement, as well as fluctuations, over the un-ligated model. This may be attributed to a small increase in structural rigidity of the EF hand loop binding sites when EGCg is bound nearby.

Table 6.16 shows the calcium coordination for the ff12SB simulations of between 8 and 9, in the range predicted by the literature of 7-9 and confirmed by the previous simulation results. Unlike the results shown in Table 6.3, there is a slightly larger range of coordination for the site 1 calcium which is in closer agreement to the earlier cCTnC results seen in Table

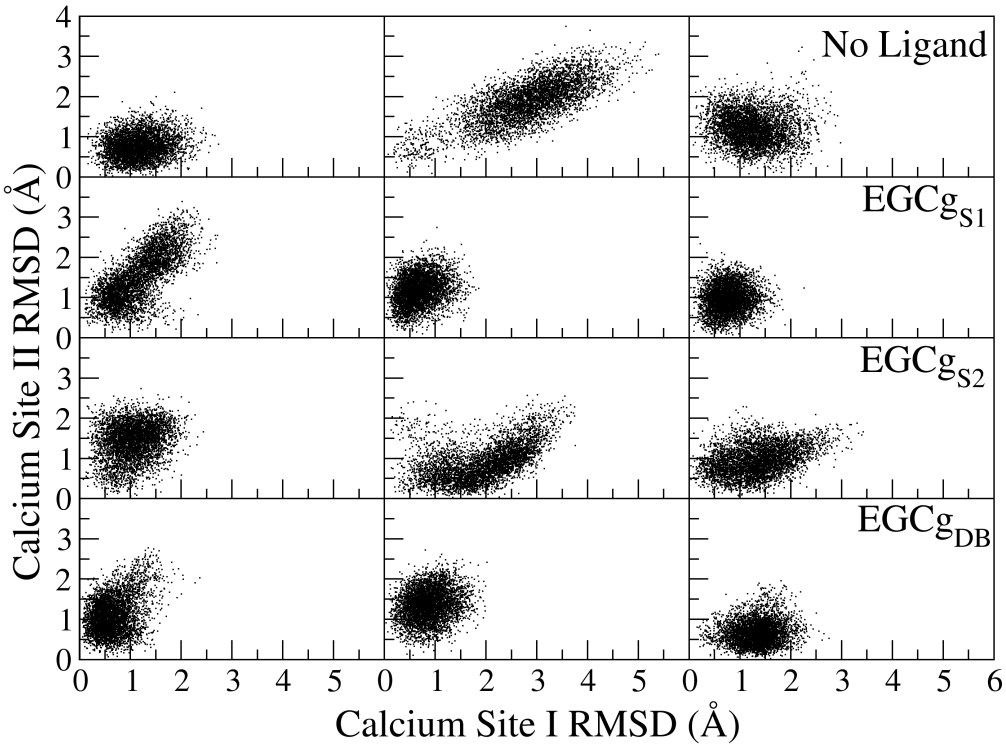

Figure 6.9: Calcium ion RMSD (Å) of Site I versus Site II from the production runs of the replica simulations.

5.5. There is also a lower coordination for site 2 in the EGCg$_{S1}$ model than is seen in the others, though again this is not a large difference overall. Despite some differences, it appears that the simulations are still in agreement in terms of oxygen coordination for both ff03 and ff12SB simulations. When compared to the calcium motion scatter plots in in Figure 6.9 there appears to be little connection between larger fluctuation and lower coordination, as shown by the small differences in the cCTnC·TnI(34-71) and EGCg$_{S2}$ replicas coordination numbers and their relatively variable motion. Table 6.16 appears to show quite consistent results over the different simulations and it appears that the relatively simple description used to model the divalent calcium ions has not been significantly affected by the change in force field from ff03 to ff12SB.

This data appears to favour the idea that drugs targeted to bind to site 1 and/or site 2 locations can have an effect on the calcium ions of these high

Table 6.16: The coordination numbers of calciums ions averaged throughout the cCTnC·TnI ff12SB MD simulations.

Model	Calcium 1	Calcium 2
cCTnC·TnI(34-71)	-	-
1	9.0	9.0
2	8.9	9.0
3	8.4	8.5
$EGCg_{S1}$	-	-
1	8.2	8.4
2	8.2	8.4
3	8.8	8.6
$EGCg_{S2}$	-	-
1	9.0	9.0
2	8.6	8.9
3	8.2	8.5
$EGCg_{DB}$	-	-
1	8.7	8.9
2	8.6	8.9
3	8.4	8.9

affinity sites, as shown in the earlier section, though this is not as distinct as was shown previously.

6.3.6 Comparison to Experimental Findings

With the results gathered from the simulations it becomes possible to compare our results with the sparse NOE assignments used to generate the initial docked configurations. As was briefly touched upon earlier, the experimentalists had difficulty assigning NOE restraints between the EGCg/s and residues of the protein when cTnI was present in the system. In particular they were unable to confirm which leucine residue, leucine 114 or isoleucine 119, were in contact with EGCg in site 2 (they thus assigned them to both to generate the original site 2 binding). By performing the same NOE analysis as shown in Chapter 5, Section 5.3 it is possible for us to see the preference of the EGCg in each replica simulation. In this case the only EGCg NOEs found experimentally were H11/H14 (A ring), H12/H13 (C ring) and H9/H10 (B′ ring) which were considered indistin-

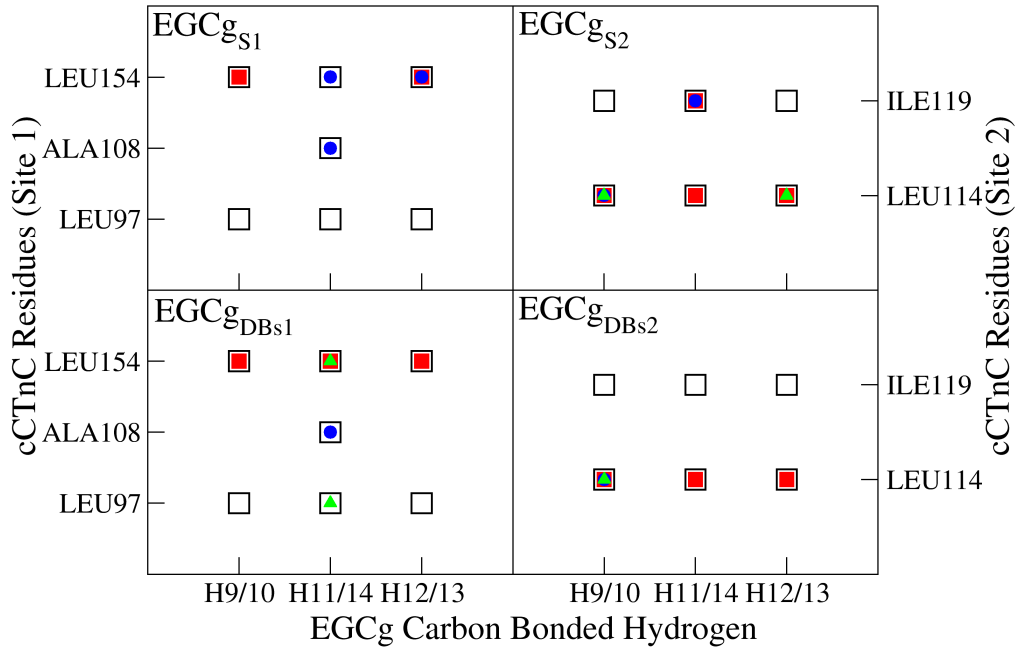

Figure 6.10: Comparison of NOE restraint assignment from experimental docking (hollow black squares) and the computational results. In each graph red squares are for replica 1, blue circles are for replica 2 and green triangles are for replica 3.

guishable from one another. These are named in the same manner as those shown in Figure 5.9. Note that the B ring H16/H17 were not included in the experimental work, nor were several other hydrogens on the benzopyran A/C ring. The results are shown in Figure 6.10 alongside the original NOEs used to generate the double bound structure.

When determining the structure from the experimental data, the restraints were all assigned values of 4.2 Å - 6.0 Å. As the calculated values almost exclusively lie within this range we have not differentiated for varying strength of the NOE. The EGCg$_{S1}$ results appear to show a complete lack of interaction with the LEU97 when in actuality there were several contacts though they were from the H16/H17 B ring. The replica 3 simulation here does not show any NOEs listed though several interactions with nearby residues are noted (see Table 6.6) including with LEU97. The EGCg$_{DBs1}$ results are quite similar to the single site case and both show a considerable preference for NOE contact with LEU154 over LEU97, though nearby H-

bond interactions indicate that the EGCg remains nearby throughout the simulations.

The Site 2 results show an interesting trend: In both cases the simulations demonstrate a considerable preference for the LEU114 NOE set over the ILE119. The $EGCg_{S2}$ results show a slight preference for the LEU114 while the $EGCg_{DBs2}$ show no contacts to the ILE119 at all. This may be an indication of why the NOEs were so difficult to assign experimentally as the binding site may shift slightly when there is an excess of EGCg in the sarcomere (likely the case if a double-binding were to occur) compared to a smaller concentration which might only transiently occupy the site 2 binding area.

Overall, We were able to shed light on this system from the less well defined set of data we started with (compared to Chapter 5). We can see that the results shown appear to indicate that EGCg binds to the cCTnC·TnI(34-71) with a low affinity and in a non-specific manner, with EGCg being able to occupy both sites but likely preferring the original site 1 area. In both single and double bound cases we see that this binding affects the calcium ion affinity, crucial to the regulation of the cardiac system.

6.4 Chapter Summary

Using protocols established and refined in Chapter 5 it was possible to use computational methods to explore and enhance the state of knowledge of the cCTnC·TnI(34-71) system when both bound and free from EGCg. Docking simulations provided an excellent basis from which sparse experimental data could be improved and expanded upon, initially by finding a suitable binding site on the protein complex for EGCg and then by exploring a possible double bound structure. The effects of single and double binding to both sites are analysed and provide an interesting insight as to why this system proved difficult for experimental methods. From the data gathered, it appears likely that the second site is of lower affinity and may be a transient binding present only when an excess of EGCg is in the system. Free energy estimates give clear indication for the changes in protein-protein binding upon mutation suggesting that calcium signalling effects could also

be a strong factor in DCM and related diseases in addition to the effects on calcium affinity. The effect of signal response is also worth further investigation as a change in cCTnC·TnI(34-71) binding is likely to modify the rate of calcium signal transmission. In all cases EGCg acted as a moderator of the interactions of the system strongly suggestive of a potential role for this or similar ligands in ameliorating the severe effects inherrent with diseased cardiac musculature.

Chapter 7

Conclusions and Perspectives

In this work a variety of computational methods have been applied to green tea catechins as well as the cardiac protein troponin C. In doing so the properties of these selected green tea catechins have been analysed in vacuo, water, chloroform and benzene solvent media. The effect of the green tea catechin EGCg on the human cardiac protein troponin C has been investigated with regards to its ability to affect calcium mobility. Simulations have provided atomic scale examination of these chemical and biological systems and their deficiencies and limitations have also been discussed. This Chapter briefly summarises of the outcomes of the previous Chapters as well as giving thought to further investigation of these systems using the work detailed here as a foundation.

Structural Properties of Catechins

Computational techniques at various levels of accuracy have been carried out to elucidate the interactions of four selected green tea catechins (EGCg, EGCmg, ECg and EGC). In doing so a complete picture of the interplay of intra-molecular and inter-molecular interactions in gas phase as well as water and chloroform environments has been achieved, giving new insights into the configurational landscape of these systems.

Classical molecular dynamics, ably aided by the use of metadynamics to accelerate rare events, has allowed us to quantify the interactions of the selected green tea catechins with the inclusion of explicit solvent media

including water and hydrophobic chloroform. We were able to construct a force field modification to the Amber GAFF file which adequately reproduces the results derived from DFT calculations. This allowed us to perform MD and metadynamics to explore the potential energy surface of these molecules using selected molecular torsions to describe the systems. We also compared these catechins with currently used calcium sensitisers Levosimendan and EMD 57033 to analyse their similarities and find the catechins possess a more hydrophilic character than these drugs. This also allowed us to further interrogate these molecules bound to biological systems later on in Chapter 5.

Electronic Properties of Catechins

The claimed antioxidant properties of green tea catechins have led to their use as a health supplement for a variety of conditions. Using DFT, Hartree Fock and Møller Plesset calculations, this work has expanded the current knowledge of the electronic properties of the EGCg and EGC catechins in various radical forms. In addition, we have paid particular attention to the current deficiencies in the modelling these systems using these methods.

We show that the galloyl B′ ring present in EGCg has little to no effect on the ionisation potential or bond dissociation enthalpies of the molecule, leading us to conclude that EGCg and EGC have very similar mechanisms. Overall we found that BDE, which characterises the HAT mechanism, is smaller than IP for SET or H^+ transfer for SPLET in both catechins. It would therefore be of interest to further investigate the HAT mechanism, for example with molecular dynamics, as this may provide more information on the radical processes involved with these molecules. The effect of solvent dielectric has also been tested and shown to be an important consideration when evaluating charged systems such as the cationic and anionic states of the catechins, though has a very minor effect on bond dissociation enthalpy. The solvent medium does considerably affect the ionisation potential as well as proton transfer when considering a water or hydrophobic medium (such as in the case of biological membranes). We find ongoing issues with the treatment of spin density in the cationic system, notably in the gas

phase but present in any system lacking a large external dielectric, using current methods which make determination of other molecular properties problematic in this case.

Interactions of EGCg and the Cardiac Troponin C Protein

Classical molecular dynamics of the protein-ligand system cCTnC·EGCg were carried out using the parametrisation generated in the earlier catechin work to model EGCg. We were able to show that computational methods could reproduce, to a good degree of accuracy, the experimental work on this protein, notably several major NOE restraints used to fit the experimental structure. In addition, our simulations were then able to give a detailed account of the interaction network that EGCg forms when in complex with cCTnC as well as other similar drugs which may be used to target the hydrophobic cleft, including EMD 57033. We used metadynamics to explore this protein cleft and show that EGCg is able to bind in different positions based on its ring configuration whilst remaining bound to the protein as well as to provide possible transition pathways for dissociation of the catechin from the complex. We also give an estimate of the free energy of binding based upon MM/PBSA results which should be useful for comparative work with similar classes of drugs designed to target this system.

Interactions of EGCg and the Cardiac Troponin C - Troponin I Complex

Using the docking protocols outlined in our earlier work we were able to locate an alternative binding site for EGCg in the case where the hydrophobic cleft of cCTnC is occupied by the cTnI(34-71) anchoring fragment, located at the interface of the E & H helices of cCTnC. With the aid of MD and MM/PBSA protocols established in Chapter 5 we have shown that while this binding is weaker than to the hydrophobic cleft, it still has an effect on the calcium mobility in the system and may act as a calcium modulater,

though it is unlikely, in itself, to solve the problems associated with cardiac malfunction. We therefore identify this site as of particular interest for further study for drugs which indirectly alter the calcium cycle of the heart.

In combination with NMR experimental work[190] on this system, we were then able to identify and test a second binding site for EGCg (site 2, at the interface of the F & G helices of cCTnC), as well as a potential double bound case, to the cCTnC·TnI(34-71) complex. We were able to use MD simulations to clarify the NOE data seen experimentally and find that EGCg likely binds non-specifically to these sites, with relatively low affinity and may only transiently bind to site 2. In this case the effect on calcium mobility was less noticeable. Nevertheless it is recommneded that both sites be persued by those seeking to alter the state of calcium binding in the cardiac sarcomere.

Ongoing Work and Future Perspectives

DFT calculations have allowed us to produce robust and useful force field modifications for the green tea catechins EGC, ECg, EGCmg and EGCg for further work with biological systems. The work in this thesis has focussed on the commonly stated 'most efficacious' of these molecules, EGCg, however our findings from extensive electronic calculations indicate that it would also be prudent to extend this investigation in the future to EGC which we have shown to be at least as effective in terms of its antioxidant related properties as EGCg, though it is less abundant than EGCg in green tea. Given the less bulky size of the smaller EGC molecule, coupled with similar energetic processes (IP, BDE) to EGCg, it may provide a useful alternative molecule to probe the effects of dietary polyphenols and should be considered in further work on the subject.

The work in Chapter 4 outlines some issues with the current treatments of radical molecules using commonly available DFT functionals. The primary issue being that as DFT methods are based upon the property of electron density, if the method cannot accurately describe the unpaired spin density of this class of organic radicals, other results may well be erroneous. This error is difficult to quantify though appears to be exacerbated by the

poly-aromatic structures of these catechins leading to excess de-localisation of spin over the system. It would be beneficial to examine these molecules using alternative methods such as Quantum Monte Carlo, which may be able to bypass some of the drawbacks of the current DFT methodology and has been demonstrated to work for radicals.[191] This work is currently ongoing with EGC as a test system and was chosen as the method is highly computationally intensive which greatly benefits from smaller system size. It is hoped that this can be used to complement the works shown here.

The regulation of the cardiac cycle represents a broad body of work the understanding of which is crucial to ameliorating effects of heart disease, currently the largest cause of death in the developed world. The effect of EGCg on cardiac calcium mobility has been extensively discussed in this work, and our results so far indicate that it is worth further analysis of the newly assigned EGCg binding sites to cCTnC·TnI(34-71).

It may be beneficial to explore further the calcium signalling response by simulating the full Troponin protein, though this effort is currently stymied by inadequate experimental resolution of the protein in question. Alternatively, the recently developed funnel metadynamics method[110] (discussed briefly in Chapter 2) could prove to be a useful tool to probe the binding sites discovered in this work, as well as the effect of mediating waters, if any. In this way it should be possible to gain a much better estimate of the absolute binding energies of EGCg to each site as well as the effect of binding one site while the other is already occupied. As was mentioned, the similarity of EGC to EGCg may make this a plausible alternative candidate, where the less branched, bulky shape of the molecule may give it more character of the calcium sensitiser EMD 57033. It may be worth using a combination Quantum Mechanics / Molecular Mechanics (QM/MM) regime based on the recently discovered binding sites in order to benefit from the wealth of data generated by our classical work and to thoroughly interrogate these sites with catechins and other drugs capable of affecting calcium binding.

Bibliography

[1] OpenStax College. *Anatomy & Physiology*. Rice University, 2013.

[2] M Kamisago, S Sharma, S DePalma, S Solomon, P Sharma, B McDonough, L Smoot, M Mullen, P Woolf, E Wigle, JG Seidman, J Jarcho, L Shapiro, and C Seidman. Mutations in sarcomere protein genes as a cause of dilated cardiomyopathy. *New Engl. J. Med.*, 343(23):1688–1696, 2000.

[3] A Laio and F L Gervasio. Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science. *Rep. Prog. Phys.*, 71:126601, 2008.

[4] S Takeda, A Yamashita, K Maeda, and Y Maeda. Structure of the core domain of human cardiac troponin in the ca^{2+} saturated form. *Nature*, 424:35–41, 2003.

[5] X Wang, M X Li, L Spyracopoulos, N Beier, M Chandra, R J Solaro, and B D Sykes. Structure of the C-domain of human cardiac troponin C in complex with the Ca^{2+} sensitizing drug EMD 57033. *J. Biol. Chem.*, 276(27):25456–66, 2001.

[6] I Robertson, M Li, and B D Sykes. Solution structure of human cardiac troponin C in complex with the green tea polyphenol, (-)-epigallocatechin 3-gallate. *J. Biol. Chem.*, 284:23012–23022, 2009.

[7] A. MacFarlane and I. MacFarlane. *The Empire of Tea: The Remarkable History of the Plant That Took Over the World*. Overlook Press, 2004.

[8] DA Balentine, SA Wiseman, and LCM Bouwens. The Chemistry of Tea Flavonoids. *Crit. Rev. Food Sci. Nutr.*, 37(8):693–704, 1997.

[9] G P Dias, N Cavegn, A Nix, M C do Nascimento B, D Stangl, M S A Zainuddin, A E Nardi, P F Gardino, and S Thuret. The role of dietary polyphenols on adult hippocampal neurogenesis: Molecular mechanisms and behavioural effects on depression and anxiety. *Oxid. Med. Cell. Longev.*, 2012.

[10] B Mirza, H Ikram, S Bilgrami, D J Haleem, and M A Haleem. Neurochemical and behavioral effects of green tea (camellia sinensis): A model study. *Pak. J. Pharm. Sci.*, 26(3):511–516, 2013.

[11] W Zhu, H Shi, Y Wei, S Wang, C Sun, Z Ding, and L Lu. Green Tea Polyphenols produce Antidepressant-like Effects in Adult Mice. *Pharmacol. Res.*, 65(1):74–80, 2012.

[12] M Nakayama, K Suzuki, M Toda, S Okubo, Y Hara, and T Shimamura. Inhibition of the infectivity of influenza virus by tea polyphenols. *Antiviral Res.*, 21:289–299, 1993.

[13] A Hamza and C Zhan. How can (-)-epigallocatechin gallate from green tea prevent HIV-1 infection? *J. Phys. Chem. B*, 110:2910–2917, 2006.

[14] A Yamaguci, M Honda, H Ikigai, Y Hara, and T Shimamura. Inhibitory effects of (-)-epigallocatechin gallate on the life cycle of human immunodeficiency virus type 1 (HIV-1). *Antiviral Research*, 53:19–34, 2002.

[15] S Wolfram, Y Wang, and F Thielecke. Anti-obesity effects of green tea: from bedside to bench. *Mol. Nutr. Food Res.*, 50:176–187, 2006.

[16] D E. Ehrnhoefer, J Bieschke, A Boeddrich, M Herbst, L Masino, R Lurz, S Engemann, A Pastore, and E E Wanker. EGCg redirects amyloidogenic polypeptides into unstructured, off-pathway oligomers. *Nat. Struct. Mol. Biol.*, 15(6):558–566, 2008.

[17] J Bieschke, J Russ, R P Friedrich, D E Ehrnhoefer, H Wobst, K Neugebauer, and E E Wanker. EGCg remodels mature α-synuclein

and amyloid-β fibrils and reduces cellular toxicity. *Proc. Natl. Acad. Sci. U. S. A.*, 107(17):7710–7715, 2010.

[18] J L Bushman. Green tea and cancer in humans: a review of the literature. *Nutr. Cancer*, 31:151–159, 1998.

[19] K Rathore, S Choudhary, A Odoi, and H R Wang. Green tea catechin intervention of reactive oxygen species-mediated ERK pathway activation and chronically induced breast cell carcinogenesis. *J. Carcinog.*, 33(1):174–183, 2012.

[20] S F Kostin, D E McDonald, and D W McFadden. Inhibitory effects of (-)-epigallocatechin-3-gallate and pterostilbene on pancreatic cancer growth in vitro. *J. Surg. Res.*, 177(2):255 – 262, 2012.

[21] F De Amicis, A Perri, D Vizza, A Russo, M L Panno, D Bonofiglio, C Giordano, L Mauro, S Aquila, D Tramontano, and S Andò. Epigallocatechin gallate inhibits growth and epithelial-to-mesenchymal transition in human thyroid carcinoma cell lines. *J. Cell. Physio.*, 228(10):2054–2062, 2013.

[22] M G L Hertog, E J M Feskens, P C H Hollman, M B Katan, and D Kromhout. Dietary antioxidant flavonoids and risk of coronary heart disease: the zutphen elderly study. *Lancet*, 339:1523–1526, 1992.

[23] R Riemersma, C Rice-Evans, R Tyrrell, M Clifford, and M Lean. Tea flavonoids and cardiovasular health. *Q. J. Med.*, 94:277, 2001.

[24] M Hodgson. Tea flavonoids and cardiovascular disease. *Asia Pac. J. Clin. Nutr.*, 17:288–290, 2008.

[25] B N Singh, S Shankar, and R K Srivastava. Green tea catechin, epigallocatechin-3-gallate (EGCg): Mechanisms, perspectives and clinical applications. *Biochem. Pharm.*, 82(12):1807–1821, 2011.

[26] T Zhang, J Zhang, P Derreumaux, and Y Mu. Molecular mechanism of the inhibition of EGCG on the Alzheimer Aβ(1-42) dimer. *J. Phys. Chem. B*, 117(15):3993–4002, 2013.

[27] World Health Organisation. Global burden of disease, 2008.

[28] N Tadano, C Du, F Yumoto, S Morimoto, M Ohta, M Xie, K Nagata, D Zhan, Q Lu, Y Miwa, F Takahashi-Yanaga, M Tanokura, I Ohtsuki, and T Sasaguri. Biological actions of green tea catechins on cardiac troponin C. *Br. J. Pharmacol.*, 161(5):1034–1043, 2010.

[29] W W Wong, J H Gerson, P A Rubenstein, and E Reisler. Thin filament regulation and ionic interactions between the N-terminal region in actin and troponin. *Biophys. J.*, 83(5):2726 – 2732, 2002.

[30] F Yanaga, S Morimoto, and I Ohtsuki. Ca^{2+} sensitization and potentiation of the maximum level of myofibrillar atpase activity caused by mutations of troponin T found in familial hypertrophic cardiomyopathy. *J. Biol. Chem.*, 274(13):8806–8812, 1999.

[31] O K Baryshnikova, I M Robertson, P Mercier, and B D Sykes. The dilated cardiomyopathy G159D mutation in cardiac troponin C weakens the anchoring interaction with troponin I. *Biochemistry*, 47(41):10950–10960, 2008.

[32] A. F. Huxley and R. Niedergerke. Structural changes in muscle during contraction: Interference microscopy of living muscle fibres. *Nature*, 173(4412):971–973, 1954.

[33] H. Huxley and J. Hanson. Changes in the cross-striations of muscle during contraction and stretch and their structural interpretation. *Nature*, 173(4412):973–976, 1954.

[34] A Fabiato. Calcium-induced release of calcium from the cardiac sarcoplasmic reticulum. *Am. J. Physiol.*, 245:C1–C14, 1983.

[35] L L Brunton. Excitation-contraction coupling and cardiac contractile force. *Trends Pharmacol. Sci.*, 13:297, 1992.

[36] H Cheng, W J Lederer, and M B Cannell. Calcium sparks: elementary events underlying excitation-contraction coupling in heart muscle. *Science (New York, N.Y.)*, 262:740–744, 1993.

[37] D M Bers. Cardiac excitation-contraction coupling. *Nature*, 415:198–205, 2002.

[38] S E Boussouf and R Maytum. Role of tropomyosin isoforms in the calcium sensitivity of striated muscle thin filaments. *J. Muscle Res. Cell Motil.*, 28(1):49–58, 2007.

[39] M Leon and L Cohen. Hypertrophic cardiomyopathy. *Disease-a-Month*, 28(2):1 – 87, 1981.

[40] E Wigle, Z Sasson, M Henderson, T Ruddy, J Fulop, H Rakowski, and W Williams. Hypertrophic cardiomyopathy. the importance of the site and the extent of hypertrophy. a review. *Prog. Cardiovasc. Dis.*, 28(1):1 – 83, 1985.

[41] T Sorsa, P Pollesello, and R Solaro. The contractile apparatus as a target for drugs against heart failure: Interaction of levosimendan, a calcium sensitiser, with cardiac troponin c. *Mol. Cell. Biochem.*, 266(1-2):87–107, 2004.

[42] T Sorsa, S Heikkinen, M Abbott, E Abusamhadneh, T Laakso, C Tilgmann, R Serimaa, A Annila, P Rosevear, T Drakenberg, P Pollesello, and I Kilpeläinen. Binding of Levosimendan, a Calcium Sensitizer, to Cardiac Troponin C. *J. Biol. Chem.*, 276(12):9337–9343, 2001.

[43] T Sorsa, P Pollesello, P Rosevear, T Drakenberg, and I Kilpeläinen. Stereoselective binding of levosimendan to cardiac troponin C causes Ca^{2+}-sensitization. *Eur. J. Pharmacol.*, 486(1):1–8, 2004.

[44] M Li, I Robertson, and B Sykes. Interaction of cardiac troponin with cardiotonic drugs: A structural perspective. *Biochem. Biophys. Res. Comm.*, 369(1):88–99, 2008.

[45] R Solaro, P Bousquet, and J Johnson. Stimulation of cardiac myofilament force, ATPase activity and troponin C Ca^{2+} binding by bepridil. *J. Pharmacol. Exp. Ther.*, 238(2):502–507, 1986.

[46] Q Kleerekoper, W Liu, D Choi, and J A Putkey. Identification of binding sites for bepridil and trifluoperazine on cardiac troponin C. *J. Biol. Chem.*, 273(14):8153–8160, 1998.

[47] L MacLachlan, D Reid, R Mitchell, C Salter, and S Smith. Binding of a calcium sensitizer, bepridil, to cardiac troponin C. A fluorescence stopped-flow kinetic, circular dichroism, and proton nuclear magnetic resonance study. *J. Biol. Chem.*, 265(17):9764–9770, 1990.

[48] J Varughese, T Baxley, J Chalovich, and Y Li. A computational and experimental approach to investigate bepridil binding with cardiac troponin. *J Phys. Chem. B*, 115(10):2392–2400, 2011.

[49] J Varguhese and Y Li. Molecular dynamics and docking studies on cardiac troponin C. *J. Biomol. Struct. Dyn.*, 29(1):123–135, 2011.

[50] J Herzig, V Tkachuk, G Baldenkov, I Feoktistov, M Menshikov, and U Quast. Calmodulin and troponin C as targets for drug action. *Biomed. Biochim. Acta.*, 46(8-9):S440–S443, 1987.

[51] N Kurebayashi and Y Ogawa. Increase by trifluoperazine in calcium sensitivity of myofibrils in a skinned fibre from frog skeletal muscle. *J. Physiol.*, 403:407–424, 1988.

[52] R Levin and B Weiss. Specificity of the binding of trifluoperazine to the calcium-dependent activator of phosphodiesterase and to a series of other calcium-binding proteins. *Biochim. Biophys. Acta.*, 540(2):197–204, 1978.

[53] S Lipscomb, L C Preston, P Robinson, C S Redwood, I P Mulligan, and C C Ashley. Effects of troponin C isoform on the action of the cardiotonic agent EMD 57033. *Biochem. J.*, 388:905–912, 2005.

[54] B Pan and R Johnson. Interaction of cardiotonic thiadiazinone derivatives with cardiac troponin C. *J. Biol. Chem.*, 271(2):817–823, 1996.

[55] M Li, L Spyracopoulos, N Beier, J Putkey, and B D Sykes. Interaction of cardiac troponin C with Ca^{2+} sensitizer EMD 57033 and cardiac troponin I inhibitory peptide. *Biochemistry*, 39(30):8782–8790, 2000.

[56] R M Tufty and R H Kretsinger. Troponin and parvalbumin calcium binding regions predicted in myosin light chain and T4 lysozyme. *Science*, 187(4172):167–169, 1975.

[57] S P Robertson, J D Johnson, and J Potter. The time-course of Ca^{2+} exchange with calmodulin, troponin, parvalbumin, and myosin in response to transient increases in Ca^{2+}. *Biophys. J.*, 34(3):559–69, June 1981.

[58] O Herzberg and M James. Common structural framework of the two calcium/magnesium binding loops of troponin C and other calcium binding proteins. *Biochemistry*, 24(20):5298–5302, 1985.

[59] J-M François, C Gerday, F Prendergast, and J Potter. Determination of the Ca^{2+} and Mg^{2+} affinity constants of troponin C from eel skeletal muscle and positioning of the single tryptophan in the primary structure. *J Muscle Res. Cell Motil.*, 14(6):585–593, December 1993.

[60] S Cates, M Berry, E Ho, Q Li, J Potter, and G Phillips. Metal-ion affinity and specificity in EF-hand proteins: coordination geometry and domain plasticity in parvalbumin. *Structure*, 7(10):1269–1278, October 1999.

[61] E Blanchard and R Solaro. Inhibition of the activation and troponin calcium binding of dog cardiac myofibrils by acidic pH. *Circ. Res.*, 55(3):382–391, 1984.

[62] *Flexible docking of the fragment of the troponin I to the troponin C*, Granada, Spain, 2014.

[63] D Lindhout and B Sykes. Structure and dynamics of the C-domain of human cardiac troponin C in complex with the inhibitory region of human cardiac troponin I. *J. Biol. Chem.*, 278(29):27024–27034, 2003.

[64] J Varguhese, J Chalovich, and Y Li. Molecular dynamics studies on troponin (TnI-TnT-TnC) complexes: insight into the regulation of muscle contraction. *J. Biomol. Struct. Dyn.*, 28(2):159–174, 2010.

[65] D Paul, E Morris, R Kensler, and J Squire. Structure and Orientation of Troponin in the Thin Filament. *J. Biol. Chem.*, 284(22):15007–15015, 2009.

[66] A Leach. *Molecular Modelling - Principles and Applications*. Prentice Hall Publications, 2nd edition edition, 2001.

[67] D C Young. *Computational Chemistry: A Practical Guide for Applying Techniques to Real-World Problems*, volume 9. 2001.

[68] A Szabo and N S Ostlund. *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*. 1996.

[69] A Montoya, TN Truong, and AF Sarofim. Spin contamination in Hartree-Fock and density functional theory wavefunctions in modeling of adsorption on graphite. *J. Phys. Chem. A*, pages 6108–6110, 2000.

[70] C Møller and M S Plesset. Note on an approximation treatment for many-electron systems. *Phys. Rev.*, 46:618–622, 1934.

[71] R G Parr and W Yang. *Density-Functional Theory of Atoms and Molecules*, volume 16. 1989.

[72] J Hutter, M Iannuzzi, F Schiffmann, and J VandeVondele. cp2k: atomistic simulations of condensed matter systems. *WIRES Comput. Mol. Sci.*, 4(1):15–25, 2014.

[73] P Hohenberg and W Kohn. Inhomogeneous electron gas. *Phys. Rev.*, 136:B864–B871, 1964.

[74] W Kohn and L J Sham. Self-consistent equations including exchange and correlation effects. *Phys. Rev.*, 140:A1133–A1138, 1965.

[75] J P Perdew. Density-functional approximation for the correlation energy of the inhomogeneous electron gas. *Phys. Rev. B*, 33:8822–24, 1986.

[76] A D Becke. Density-functional thermochemistry. IV. a new dynamical correlation functional and implications for exact-exchange mixing. *J. Chem. Phys.*, 104:1040–46, 1996.

[77] C Lee, W Yang, and R G Parr. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B*, 37:785–789, 1988.

[78] C Adamo and V Barone. Toward reliable density functional methods without adjustable parameters: The PBE0 model. *J. Chem. Phys.*, 110:6158, 1999.

[79] A D Becke. Density-functional thermochemistry. III. the role of exact exchange. *J. Chem. Phys.*, 98:5648–52, 1993.

[80] M E Tuckerman and G J Martyna. Understanding modern molecular dynamics: Techniques and applications. *J. Phys. Chem. B*, 104:159, 2000.

[81] J P Perdew and A Zunger. Self-interaction correction to density-functional approximations for many-electron systems. *Phys. Rev. B*, 23:5048–5079, 1981.

[82] T Yanai, D P. Tew, and N C. Handy. A new hybrid exchange-correlation functional using the Coulomb-attenuating method (CAM-B3LYP). *Chem. Phys. Lett.*, 393:51–57, 2004.

[83] Y Andersson, D Langreth, and B Lundqvist. van der waals interactions in density-functional theory. *Phys. Rev. Lett.*, 76:102–105, Jan 1996.

[84] von Lilienfeld.

[85] O von Lilienfeld, I Tavernelli, U Rothlisberger, and D Sebastiani. Performance of optimized atom-centered potentials for weakly bonded systems using density functional theory. *Phys. Rev. B*, 71(19), 2005.

[86] S Grimme. Accurate description of van der Waals complexes by density functional theory including empirical corrections. *J. Comput. Chem.*, 25(12):1463–1473, 2004.

[87] S Grimme. Semiempirical GGA-type density functional constructed with a long-range dispersion correction. *J. Comput. Chem.*, 27:1787–1799, 2006.

[88] S Grimme, J Antony, S Ehrlich, and H Krieg. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.*, 132, 2010.

[89] E Johnson and G DiLabio. Structure and binding energies in van der Waals dimers: Comparison between density functional theory and correlated ab initio methods. *Chem. Phys. Lett.*, 419(4):333–339, 2006.

[90] J Ramalho, J Gomes, and F Illas. Accounting for van der Waals interactions between adsorbates and surfaces in density functional theory based calculations: selected examples. *RSC Adv.*, 3(32):13085–13100, 2013.

[91] M Sardo, R Siegel, S Santos, J Rocha, J Gomes, and L Mafra. Combining Multinuclear High-Resolution Solid-State MAS NMR and Computational Methods for Resonance Assignment of Glutathione Tripeptide. *J. Phys. Chem. A*, 116(25):6711–6719, 2012.

[92] JP Ryckaert, G Ciccotti, and HJC Berendsen. Numerical integration of the cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes. *J. Comput. Phys.*, 23:327–341, 1977.

[93] D Frenkel and B Smit. *Understanding Molecular Simulation From Algorithms to Applications*. Academic Press, 2nd edition edition, 2002.

[94] J Ponder and D Case. Force fields for protein simulations. *Adv. Prot. Chem.*, 66:27–85, 2003.

[95] D A Case, T A Darden, T E Cheatham, C L Simmerling, J Wang, R E Duke, R Luo, R C Walker, W Zhang, K M Merz, B Roberts, S Hayik, A Roitberg, G Seabra, J Swails, A W Goetz, I Kolossváry, K F Wong, F Paesani, J Vanicek, R M Wolf, J Liu, X Wu, S R Brozell, T Steinbrecher, H Gohlke, Q Cai, X Ye, J Wang, M J Hsieh, G Cui, D R Roe, D H Mathews, M G Seetin, R Salomon-Ferrer, C Sagui, V Babin, T Luchko, S Gusarov, A Kovalenko, and P A Kollman. AMBER 12, 2012.

[96] A Toukmaji, C Sagui, J Board, and T Darden. Efficient particle-mesh Ewald based approach to fixed and induced dipolar interactions. *J. Chem. Phys*, 113:10913, 2000.

[97] H J C Berendsen, J P M Postma, W F van Gunsteren, A DiNola, and J R Haak. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.*, 81:3684–3690, 1984.

[98] D S Cerutti, R Duke, P L Freddolino, H Fan, and T P Lybrand. Vulnerability in popular molecular dynamics packages concerning langevin and anderson dynamics. *J. Chem. Theory Comput.*, 4(10):1669–1680, 2008.

[99] G S Grest and K Kremer. Molecular-dynamics simulation for polymers in the presence of a heat bath. *Phys. Rev. A*, 33(5):3628–3631, 1986.

[100] W Cornell, P Cieplak, C Bayly, I Gould, K Merz, D Ferguson, D Spellmeyer, T Fox, J Caldwell, and P Kollman. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.*, 117:5179–5197, 1996.

[101] S Weiner, P Kollman, D Case, U Singh, and C Ghio. A new force field for molecular mechanical simulation of nucleic acids and proteins. *J. Am. Chem. Soc.*, 106:765–784, 1984.

[102] S Weiner, P Kollman, Dzung T Nguyen, and D Case. An all atom force field for simulations of proteins and nucleic acids. *J. Comput. Chem.*, 7:230–252, 1986.

[103] V Hornak, R Abel, A Okur, B Strockbine, A Roitberg, and C Simmerling. Comparison of multiple amber force fields and development of improved protein backbone parameters. *Proteins*, 65:712–725, 2006.

[104] K Lindorff-Larsen, S Piana, K Palmo, P Maragakis, J L Klepeis, R O Dror, and D E Shaw. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins*, 78(8):1950–8, 2010.

[105] Y Duan, C Wu, S Chowdhury, M C Lee, G Xiong, W Zhang, R Yang, P Cieplak, R Luo, T Lee, J W Caldwell, J Wang, and P A Kollman.

A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.*, 24:1999–2012, 2003.

[106] D Shaw, P Maragakis, K Lindorff-Larsen, S Piana, R Dror, M Eastwood, J Bank, J Jumper, J Salmon, Y Shan, and W Wriggers. Atomic-level characterization of the structural dynamics of proteins. *Science*, 330(6002):341–346, 2010.

[107] A Laio and M Parrinello. Escaping free energy minima. *Proc. Natl. Acad. Sci. U. S. A.*, 99:12562–12566, 2002.

[108] M Bonomi, D Branduardi, G Bussi, C Camilloni, D Provasi, P Raiteri, D Donadio, F Marinelli, F Pietrucci, R A Broglia, and M Parrinello. PLUMED: A portable plugin for free-energy calculations with molecular dynamics. *Comput. Phys. Commun.*, 180:1961–1972, 2009.

[109] A Barducci, G Bussi, and M Parrinello. Well-tempered metadynamics: A smoothly converging and tunable free-energy method. *Phys. Rev. Lett.*, 100:020603, 2008.

[110] V Limongelli, M Bonomi, and M Parrinello. Funnel metadynamics as accurate binding free-energy method. *Proc. Natl. Acad. Sci. U.S.A.*, 110:6358–63, 2013.

[111] W Wang and P A Kollman. Free energy calculations on dimer stability of the HIV protease using molecular dynamics and a continuum solvent model. *J. Mol. Biol.*, 303:567–582, 2000.

[112] A Weis, K Katebzadeh, P Söderhjelm, I Nilsson, and U Ryde. Ligand affinities predicted with the mm/pbsa method: Dependence on the simulation method and the force field. *J. Med. Chem.*, 49(22):6596–6606, 2006. PMID: 17064078.

[113] D L Beveridge and F M DiCapua. Free energy via molecular simulation. applications to chemical and biomolecular systems. *Annu. Rev. Biophys. Bio.*, 18:431–492, 1989.

[114] H Gohlke and G Klebe. Approaches to the description and prediction of the binding affinity of small-molecule ligands to macromolecular receptors. *Angew. Chem. Int. Ed.*, 41(15):2644–2676, 2002.

[115] A R Brice and B N Dominy. Analyzing the robustness of the MM/PBSA free energy calculation method: Application to DNA conformational transitions. *J. Comput. Chem.*, 32(7):1431–1440, 2011.

[116] P A Kollman, I Massova, C Reyes, B Kuhn, S Huo, L Chong, M Lee, T Lee, Y Duan, W Wang, O Donini, P Cieplak, J Srinivasan, D A Case, and T E Cheatham. Calculating structures and free energies of complex molecules: Combining molecular mechanics and continuum models. *Acc. Chem. Res.*, 33(12):889–897, 2000. PMID: 11123888.

[117] H Gohlke and DA Case. Converging free energy estimates: MM–PB(GB)SA studies on the protein–protein complex Ras–Raf. *J. Comput. Chem.*, pages 238–250, 2004.

[118] K A Sharp and B Honig. Electrostatic interactions in macromolecules: theory and applications. *Annu. Rev. Biophys. Biophys. Chem.*, 19:301–332, 1990.

[119] T Hou, J Wang, Y Li, and W Wang. Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations. *J. Chem. Info. Model.*, 51(1):69–82, 2011.

[120] I Klapper, R Hagstrom, R Fine, K Sharp, and B Honig. Focusing of electric fields in the active site of Cu-Zn superoxide dismutase: Effects of ionic strength and amino-acid modification. *Proteins*, 1:47–59, 1986.

[121] M K Gilson and B Honig. Calculation of the total electrostatic energy of a macromolecular system: solvation energies, binding energies, and conformational analysis. *Proteins*, 4:7–18, 1988.

[122] A Nicholls and B Honig. A Rapid Finite-Difference Algorithm, Utilizing Successive over-Relaxation to Solve the Poisson-Boltzmann Equation. *J. Comput. Chem.*, 12:435–445, 1991.

[123] R Luo, L David, and M K Gilson. Accelerated Poisson-Boltzmann calculations for static and dynamic systems. *J. Comput. Chem.*, 23:1244–1253, 2002.

[124] B Lee and F M Richards. The interpretation of protein structures: estimation of static accessibility. *J. Mol. Biol.*, 55:379–400, 1971.

[125] A Shrake and J A Rupley. Environment and exposure to solvent of protein atoms. Lysozyme and insulin. *J. Mol. Biol.*, 79(2):351–71, 1973.

[126] R Taylor, P Jewsbury, and J Essex. A review of protein-small molecule docking methods. *J. Comp. Mole. Des.*, 16:151–166, 2002.

[127] G Morris, D Goodsell, R Halliday, R Huey, W Hart, R Belew, and A Olson. Automated docking using a lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.*, 19(14):1639–1662, 1998.

[128] M Sanner. Python: A programming language for software integration and development. *J. Mol. Graphics Modell.*, 17:57–61, 1999.

[129] G Morris, D Goodsell, R Halliday, R Huey, W Hart, R Belew, and A Olson. Automated docking using a lamarckian genetic algorithm and an empirical binding free energy function. *J. Comp. Chem.*, 19(14):1639, 1998.

[130] K R Landis-Piwowar, C Huo, D Chen, V Milacic, G Shi, T H Chan, and Q P Dou. A novel prodrug of the green tea polyphenol (-)-epigallocatechin-3-gallate as a potential anticancer agent. *Cancer Res.*, 67(9):4303–10, 2007.

[131] J Kanwar, M Taskeen, I Mohammad, C Huo, T H Chan, and Q P Dou. Recent advances on tea polyphenols. *Front. Biosci. (Elite Ed.)*, pages 111–131, 2012.

[132] M J Frisch, G W Trucks, H B Schlegel, G E Scuseria, M A Robb, J R Cheeseman, G Scalmani, V Barone, B Mennucci, G A Petersson, H Nakatsuji, M Caricato, X Li, H P Hratchian, A F Izmaylov,

J Bloino, G Zheng, J L Sonnenberg, M Hada, M Ehara, K Toyota, R Fukuda, J Hasegawa, M Ishida, T Nakajima, Y Honda, O Kitao, H Nakai, T Vreven, J A Montgomery, Jr., J E Peralta, F Ogliaro, M Bearpark, J J Heyd, E Brothers, K N Kudin, V N Staroverov, R Kobayashi, J Normand, K Raghavachari, A Rendell, J C Burant, S S Iyengar, J Tomasi, M Cossi, N Rega, J M Millam, M Klene, J E Knox, J B Cross, V Bakken, C Adamo, J Jaramillo, R Gomperts, R E Stratmann, O Yazyev, A J Austin, R Cammi, C Pomelli, J W Ochterski, R L Martin, K Morokuma, V G Zakrzewski, G A Voth, P Salvador, J J Dannenberg, S Dapprich, A D Daniels, Ö Farkas, J B Foresman, J V Ortiz, J Cioslowski, and D J Fox. Gaussian 09 Revision A.1. Gaussian Inc. Wallingford CT 2009.

[133] M Head-Gordon, J A Pople, and M J Frisch. MP2 energy evaluation by direct methods. *Chem. Phys. Lett.*, 153:503–506, 1988.

[134] S Sæbø and J Almlöf. Avoiding the integral storage bottleneck in LCAO calculations of electron correlation. *Chem. Phys. Lett.*, 154:83–89, 1989.

[135] M J Frisch, M Head-Gordon, and J A Pople. A direct MP2 gradient method. *Chem. Phys. Lett.*, 166:275–280, 1990.

[136] M J Frisch, M Head-Gordon, and J A Pople. Semi-direct algorithms for the MP2 energy and gradient. *Chem. Phys. Lett.*, 166:281–289, 1990.

[137] M Head-Gordon and T Head-Gordon. Analytic MP2 frequencies without fifth-order storage. Theory and application to bifurcated hydrogen bonds in the water hexamer. *Chem. Phys. Lett.*, 220:122–128, 1994.

[138] J M Wang, R M Wolf, J W Caldwell, P A Kollman, and D A Case. Development and testing of a general amber force field. *J. Comput. Chem.*, 25(9):1157–1174, 2004.

[139] D Case, T Cheatham, T Darden, H Gohlke, R Luo, K Merz, A Onufriev, C Simmerling, B Wang, and R Woods. The Amber

biomolecular simulation programs. *J. Comput. Chem.*, 26:1668–1688, 2005.

[140] W L Jorgensen, J Chandrasekhar, J D Madura, R W Impey, and M L Klein. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, 79:926–935, 1983.

[141] S Okabe, M Suganuma, M Hayashi, E Sueoka, A Komori, and H Fujiki. Mechanisms of Growth Inhibition of Human Lung Cancer Cell Line, PC–9, by Tea Polyphenols. *Jpn. J. Cancer Res.*, 88:639–43, 1997.

[142] K Dodo, T Minato, and Y Hashimoto. Structure-Activity Relationship of Bis-Galloyl Derivatives Related to (-)-Epigallocatechin Gallate. *Chem. Pharm. Bull.*, 57(2):190–194, 2009.

[143] A M Katz and B H Lorell. Regulation of cardiac contraction and relaxation. *Circulation*, 102:IV69–IV74, 2000.

[144] M Kawai, J A Lee, and C H Orchard. Effects of the Ca^{2+} sensitizer EMD 57033 on intracellular Ca^{2+} in rat ventricular myocytes: relevance to arrhythmogenesis during positive inotropy. *Clin. Sci. (Lond)*, 99:547–554, 2000.

[145] H Senzaki, T Isoda, N Paolocci, U Ekelund, JM Hare, and DA Kass. Improved mechanoenergetics and cardiac rest and reserve function of in vivo failing heart by calcium sensitizer EMD-57033. *Circulation*, 101(9):1040–1048, 2000.

[146] M Endoh. Cardiac Ca^{2+} signaling and Ca^{2+} sensitizers. *Circ. J.*, 72(12):1915–25, 2008.

[147] Z Papp, I Édes, S Fruhwald, S G De Hert, M Salmenperä, H Leppikangas, A Mebazaa, G Landoni, E Grossini, P Caimmi, A Morelli, F Guarracino, R H G Schwinger, S Meyer, L Algotsson, B G Wikström, K Jörgensen, G Filippatos, J T Parissis, M J G González, A Parkhomenko, M B Yilmaz, M Kivikko, P Pollesello, and F Follath. Levosimendan: molecular mechanisms and clinical implications:

consensus of experts on the mechanisms of action of levosimendan. *Int. J. Cardiol.*, 159(2):82–7, 2012.

[148] M Wierzejewska and M Saldyka. Are hydrogen bonds to sulfur and oxygen different? Theoretical study of dimethylsulfide and dimethylether complexes with nitric acid. *Chem. Phys. Lett.*, pages 143–147, 2004.

[149] D Weinberg, C Gagliardi, J Hull, C Murphy, C Kent, B Westlake, A Paul, D Ess, D McCafferty, and T Meyer. Proton-Coupled Electron Transfer. *Chemical Reviews*, 112(7):4016–4093, 2012.

[150] J D Lambert and R J Elias. The antioxidant and pro-oxidant activities of green tea polyphenols: A role in cancer prevention. *Arch. Biochem. Biophys.*, 501:65–72, 2010.

[151] M Leopoldini, N Russo, and M Toscano. The molecular basis of working mechanism of natural polyphenolic antioxidants. *Food Chem.*, 125:288–306, 2011.

[152] B D Craft, A L Kerrihard, R Amarowicz, and R B Pegg. Phenol-Based Antioxidants and the In Vitro Methods Used for Their Assessment. *Compr. Rev. Food Sci. Food Safety*, 11(2):148–173, 2012.

[153] J S Wright, E R Johnson, and G A DiLabio. Predicting the activity of phenolic antioxidants: theoretical method, analysis of substituent effects, and application to major families of antioxidants. *J. Am. Chem. Soc.*, 123:1173–1183, 2001.

[154] N Q Zhu, S M Sang, T C Huang, N S Bai, C S Yang, and C T Ho. Antioxidant chemistry of green tea catechins: Oxidation products of (-)-epigallocatechin gallate and (-)-epigallocatechin with peroxidase. *J. Food Lipids*, 7:275–282, 2000.

[155] S V. Jovanovic, Y Hara, S Steenken, and M G. Simic. Antioxidant potential of gallocatechins. A pulse radiolysis and laser photolysis study. *J. Am. Chem. Soc.*, 117:9881–9888, 1995.

[156] M Leopoldini, T Marino, N Russo, and M Toscano. Antioxidant Properties of Phenolic Compounds: H-Atom versus Electron Transfer Mechanism. *J. Phys. Chem. A*, 108(22):4916–4922, 2004.

[157] AD Becke. Density-functional thermochemistry .III. the role of exact exchange. *J. Chem. Phys.*, 98:5648–5652, 1993.

[158] Y Zhao, N E Schultz, and D G Truhlar. Design of density functionals by combining the method of constraint satisfaction with parametrization for thermochemistry, thermochemical kinetics, and noncovalent interactions. *J. Chem. Theory Comput.*, 2:364–382, 2006.

[159] Y Zhao and D G Truhlar. Density functionals with broad applicability in chemistry. *Acc. Chem. Res.*, 41:157–167, 2008.

[160] M Monajjemi, M Khosravi, B Honaparvar, and F Mollaamin. Substituent and Solvent Effects on the Structural Bioactivity and Anticancer Characteristic of Catechin as a Bioactive Constituent of Green Tea. *Int. J. Quant. Chem.*, 111(2011):2771–2777, 2011.

[161] J A M Simões R M B dos Santos. Energetics of the OH bond in phenol and substituted phenols: a critical evaluation of literature data. *J. Phys. Chem. Ref. Data*, 27(3), 1998.

[162] S Valcic, A Muders, N E Jacobsen, D C Liebler, and B N Timmermann. Antioxidant chemistry of green tea catechins. Identification of products of the reaction of (-)-epigallocatechin gallate with peroxyl radicals. *Chem. Res. Toxicol.*, 12(4):382–6, 1999.

[163] S Valcic, J A Burr, B N Timmermann, and D C Liebler. Antioxidant chemistry of green tea catechins. New oxidation products of (-)-epigallocatechin gallate and (-)-epigallocatechin from their reactions with peroxyl radicals. *Chem. Res. Toxicol.*, 13(9):801–10, 2000.

[164] D Botten, G Fugallo, F Fraternali, and C Molteni. A Computational Exploration of the Interactions of the Green Tea Polyphenol (-)-Epigallocatechin 3-Gallate with Cardiac Muscle Troponin C. *PLOS ONE*, 8(7), 2013.

[165] M S Parmacek and R J Solaro. Biology of the troponin complex in cardiac myocytes. *Prog. Cardiovasc. Dis.*, 47(3):159–176, 2004.

[166] S Sia, M Li, L Spyracopoulos, S Gagne, W Liu, J Putkey, and B D Sykes. Structure of cardiac muscle troponin C unexpectedly reveals a closed regulatory domain. *J. Biol. Chem.*, 272(29):18216–18221, 1997.

[167] L Spyracopoulos, M Li, S Sia, S Gagne, M Chandra, R Solaro, and B D Sykes. Calcium-induced structural transition in the regulatory domain of human cardiac troponin C. *Biochemistry*, 36(40):12138–12146, 1997.

[168] P M Kekenes-Huskey, S Lindert, and J A McCammon. Molecular basis of calcium-sensitizing and desensitizing mutations of the human cardiac troponin c regulatory domain: A multi-scale simulation study. *PLoS Comput. Biol.*, 8(11):e1002777, 2012.

[169] S Lindert, P M Kekenes-Huskey, G Huber, L Pierce, and J A McCammon. Dynamics and Calcium Association to the N-Terminal Regulatory Domain of Human Cardiac Troponin C: A Multiscale Computational Study. *J. Phys. Chem. B*, 116(29, SI):8449–8459, 2012.

[170] P Mercier, M Li, and B D Sykes. Role of the structural domain of troponin C in muscle regulation: NMR studies of Ca^{2+} binding and subsequent interactions with regions 1-40 and 96-115 of troponin I. *Biochemistry*, 39(11):2902–2911, 2000.

[171] M Packer, J Carver, R Rodeheffer, R Ivanhoe, R DiBianco, S Zeldis, G Hendrix, W Bommer, U Elkayam, M Kukin, G Mallis, J Sollano, J Shannon, PK Tandon, and D DeMets. Effect of oral milrinone on mortality in severe chronic heart failure. *New Engl. J. Med.*, 325(21):1468–1475, 1991.

[172] Y Liou, S Kuo, and S Hsieh. Differential effects of a green tea-derived polyphenol (-)-epigallocatechin-3-gallate on the acidosis-induced decrease in the Ca^{2+} sensitivity of cardiac and skeletal muscle. *Pflugers Arch., EJP*, 456(5):787–800, 2008.

[173] O Herzberg and M James. Refined crystal structure of troponin C from turkey skeletal muscle at 2·0 Åresolution. *J. Mol. Biol.*, 203(3):761–779, October 1988.

[174] J Åqvist. Ion-water interaction potentials derived from free energy perturbation simulations. *J. Phys. Chem.*, 94(21):8021–8024, 1990.

[175] G B McGaughey, M Gagné, and A K Rappé. π-Stacking interactions. Alive and well in proteins. *J. Biol. Chem.*, 273:15458–15463, 1998.

[176] B Zagrovic and W F van Gunsteren. Comparing atomistic simulation data with the NMR experiment: How much can NOEs actually tell us? *Proteins*, 63:210–218, 2006.

[177] D Dweck, D P Reynaldo, J R Pinto, and J D Potter. A dilated cardiomyopathy troponin C mutation lowers contractile force by reducing strong myosin-actin binding. *J. Biol. Chem.*, 285(23):17371–17379, 2010.

[178] A Katz, J Glusker, S Beebe, and C Bock. Calcium Ion Coordination: A Comparison with That of Beryllium, Magnesium, and Zinc. *J. Am. Chem. Soc.*, 118(24):5752–5763, 1996.

[179] M Cates, M Teodoro, and G Phillips. Molecular mechanisms of calcium and magnesium binding to parvalbumin. *Biophys. J.*, 82(3):1133–1146, 2002.

[180] B R Miller, T D McGee, J M Swails, N Homeyer, H Gohlke, and A E Roitberg. MMPBSA.py: An efficient program for end-state free energy calculations. *J. Chem. Theory Comput.*, 8(9):3314–3321, 2012.

[181] R M Levy, L Y Zhang, E Gallicchio, and A K Felts. On the nonpolar hydration free energy of proteins: Surface area and continuum solvent models for the solute-solvent interaction energy. *J. Am. Chem. Soc.*, 125(31):9523–9530, 2003.

[182] N Singh and A Warshel. Absolute binding free energy calculations: On the accuracy of computational scoring of protein-ligand interactions. *Proteins*, 78(7):1705–1723, 2010.

[183] I S Moreira, P A Fernandes, and M J Ramos. Computational alanine scanning mutagenesis—an improved methodological approach. *J. Comput. Chem.*, 28(3):644–654, 2007.

[184] L T Chong, J W Pitera, W C Swope, and V S Pande. Comparison of computational approaches for predicting the effects of missense mutations on p53 function. *J. Mol. Graphics Model.*, 27(8):978 – 982, 2009.

[185] S Takeda, A Yamashita, K Maeda, and Y Maeda. Structure of the core domain of human cardiac troponin in the ca^{2+}-saturated form. *Nature*, 424(6944):35–41, 2003.

[186] D A Lindhout, R F Boyko, D C Corson, M X Li, and B D Sykes. The role of electrostatics in the interaction of the inhibitory region of troponin I with troponin C. *Biochemistry*, 44(45):14750–14759, 2005.

[187] C D Schwieters, J J Kuszewski, N Tjandra, and G Clore. The Xplor-NIH NMR molecular structure determination package. *J. Magn. Res.*, 160:65–73, 2003.

[188] C D Schwieters, J J Kuszewski, and G Clore. Using Xplor-NIH for NMR molecular structure determination. *Prog. Nucl. Magn. Reson. Spectrosc.*, 48:47–62, 2006.

[189] T E Cheatham and D A Case. Twenty-five years of nucleic acid simulations. *Biopolymers*, 99:969–977, 2013.

[190] I Robertson. Private communication, 2013.

[191] A Zen, B L Trout, and L Guidoni. Properties of reactive oxygen species by quantum Monte Carlo. *J. Chem. Phys.*, 141:014305, 2014.